HUMAN FACTORS IN FIELD TESTING

W. G. MATHENY
LIFE SCIENCES, INC.

PREPARED FOR

ENGINEERING PSYCHOLOGY PROGRAMS
OFFICE OF NAVAL RESEARCH

UNDER

CONTRACT N00014-67-C-0315

WORK UNIT NO. 196-067

APRIL 1969

LIFE SCIENCES, INC.

HUMAN FACTORS IN FIELD TESTING

TECHNICAL REPORT

W. G. MATHENY

PREPARED FOR

ENGINEERING PSYCHOLOGY PROGRAMS
PSYCHOLOGICAL SCIENCES DIVISION
OFFICE OF NAVAL RESEARCH

CONTRACT N00014-67-C-0315-~~P001~~

APRIL 1969

LIFE SCIENCES, INC.
7305 GRAPEVINE HWY.
FORT WORTH, TEXAS 76118

ABSTRACT

This study examined the problem of human factors field evaluation within the Navy for the purpose of formulating recommendations for improved evaluative methods and techniques. Conclusions and recommendations were drawn from review of relevant general literature and Navy documents, interviews and discussions with individuals experienced in human factors evaluation, and an examination of the design and evaluation procedures carried out during the development and test of a specific Navy aircraft system.

Conclusions and recommendations are enumerated in Section 6.0 of the report along with references to the relevant supporting Section(s) within the body of the report.

Briefly, it was concluded that human factors evaluation does not receive emphasis or support comparable to that given equipment evaluation or commensurate with the importance of the human operator to the successful functioning of the system. Much more definitive and timely information must be provided the human factors field evaluator, evaluations must be more mission oriented rather than cockpit centered, the role of the mock-up inspection needs redefinition, assignment of trained Navy human factors personnel to advise and assist the contractor during development is recommended as is assignment of contractor human factors personnel during field evaluations, close cooperation between human factors and equipment design evaluation personnel during field evaluation will greatly increase the effectiveness of the evaluation, and a short intensive training course in human factors evaluation problems and methods is recommended for Navy personnel assigned to system evaluation.

ACKNOWLEDGEMENTS

The author of this report is grateful to all those individuals and organizations who supplied information and guidance during the course of this work. Their suggestions and the sharing of their experiences has made this report possible. Those distortions and inaccuracies in summarizing their contributions must be attributed to the author's own biases and he assumes that responsibility with apologies.

Especial appreciation is expressed to:

Mr. G. S. Malecki and his colleagues Drs. James W. Miller and Marshall J. Farr of the Engineering Psychology Programs Office for guidance and technical support during the course of the study.

Commander John P. Charles, OP-701H, Office of the Chief of Naval Operations, Lt. Cmdr. Robert J. Wherry, Jr., Naval Missile Center, Dr. H. G. Moore and Mr. A. I. Rubinstein, Naval Material Command and Dr. K. F. Thomson, Mr. Roger Seltz and Lt. Richard Gaines of the Naval Air Test Center for their ideas and experiences relevant to Human Factors field evaluation within the Navy.

Messrs. R. J. Garnett, Project Engineer, J. E. Burke, Supervisor Human Factors, C. E. Sparks, Supervisor, Crew Stations Design, J. Waxler, Lead, Human Factors Engineer and all the other individuals at Ling-Temco-Vought who guided the author through the intricacies of the A-7 aircraft system.

Mr. G. I. Evans of General Dynamics/Fort Worth for sharing his experience in Human Factors field evaluation of aircraft system. They have proved most relevant and useful.

Mrs. Betty Acton of Life Sciences, Inc. for coping with the problems and details of getting all of the thoughts and ideas typed and into acceptable report form.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTD)

TABLE OF ILLUSTRATIONS

1.0 PROBLEM SUMMARY

This study was initiated in response to the expressed need for a practical and reliable means of human factors evaluation during the field testing of complex weapons systems. Its objective was the production of practical and concrete suggestions for meeting this need.

From the initial broad statement of the problem the study was narrowed to focus upon the development of human factors test and evaluation techniques for operators of aircraft systems. The decision to focus attention upon these systems was based upon deliberations as to how the study might be most effectively carried out. Aircraft systems were chosen as most suitable, particularly as initial systems for study, for the following reasons:

1. They are major systems within the Navy.

2. They are complex in that they embody complex subsystems.

3. Human factors considerations are heavily involved in their operation.

4. Human factors specifications and requirements are laid on their development, test and evaluation.

5. They involve subsystems whose test and evaluation procedures may be generalizable to other Navy systems.

In order to study the human factors test and evaluation process at first hand and to base recommendations upon knowledge of the actual workings of the RDT&E process a specific aircraft system was chosen for detailed study. The particular system chosen, the A-7A aircraft, was selected since its recency in completing the RDT&E process suggested that it was representative of current practice and personnel knowledgeable about the system were available for interview and discussion.

As the work progressed the human factors efforts for the A-7E and the P-3C aircraft were examined for particular information about analyses and timing of human factors reports.

The detailed study of the A-7A was supplemented by an extensive review of formal and informal reports and writings relevant to the problem and by interviews with human factors personnel, administrators and evaluation project personnel. A reference list of the reports found most helpful are given in the Bibliography of this report.

2.0 HUMAN FACTORS EVALUATION - A FRAME OF REFERENCE

2.1 TOTAL SYSTEM OR COMPONENT EVALUATION

In considering the field evaluation of a particular system one is likely to assume that the total system is in some way evaluated with respect to its capabilities for achieving some specified goal or mission. Under this assumption one might conclude that the field test of the system is a simple matter of determining on a "go-no go" basis whether or not the system meets the mission criteria. If such were the case, specific concern with evaluation of the human component (indeed the evaluation of any other component of the system) would seem to be unnecessary busy work when carried out in connection with a field test of the system.

The assumption that human factors evaluation is a necessary and somehow distinctive part of the field evaluation required some justification and clarification - at least to this investigator. An answer seemed to be needed to the question as to why one should be concerned with the minutiae of evaluating components and subsystems of a system if the evaluative decision is one of "accept" or "reject" the total system. Presumably if the total system successfully accomplishes its design mission we would have no interest in measuring the performance of any of its components - human or hardware. If the system meets its mission criteria we might assume that the hardware and human components are functioning so as to bring about total mission success.

As it turns out there are practical and cogent reasons why component and subsystem evaluation is necessary. However, the conclusion has been reached by this investigator that evaluations go on in this somewhat piece-meal way without a clear statement as to why they do, why it is necessary that they do and what is required in order for them to be more effective under such a system of operation. This applies particularly to evaluations of the human component of the system.

The reasons for conducting component and subsystem evaluation rather than an overall system evaluation for most systems, when examined closely, tend to bring into focus the problems of and the requirements for carrying out an adequate human component field evaluation. It is hoped that the following paragraphs help to crystallize the problem and form a basis and a rationale for the recommendations given later.

2.2.1. The Nature of the Development Process

The first important reason and necessity for being concerned with component testing during field evaluation lies in the nature of the development process itself. In theory, at the beginning of the development cycle the mission of the system is delineated in detail with criteria for successful performance clearly spelled out. Many considerations mitigate against such a clear delineation.

When a system is developed to execute a new mission or to extend the capabilities for executing a present mission, the details of the system and its performance criteria cannot be stated definitively at the outset. Generally, rather explicit overall system criteria are established to be attained through application of the present state of the equipment art or the projected state of that art. However, details of the mission and the performance criteria develop with the development of the system - or perhaps more precisely, with the development of the hardware for the system. As the iterative process of design proceeds the details of system components and their requisite individual performance criteria emerge. The overall goal or mission of the system is broken down into intermediate or secondary goals for subsystems and components of the total system. The attainment of these intermediate or secondary goals by the components and subsystems are intended to summate into attainment of the overall goal by the system.

During the development process a system is being synthesized from components chosen after an analytical exercise in which the total system requirements have been broken down into subsystem and component requirements. Actually the processes of both analysis and synthesis go on in iterative fashion throughout the development period. The central point to be made, however, is that system synthesis is attained through selection of components and subsystems which can perform in accordance with the requirements made explicit by the analysis. Components are chosen (1) whose input-output characteristics match adjacent components, (2) which perform the proper transformations on the input and (3) which perform their proper function within the required time. Component and subsystem performance summate to total system performance. One needs only to reflect on the process of synthesizing a simple audio circuit to understand how the incompatibility of one component can lead to total system failure.

When system development is viewed as a process of synthesis of compatible components it is obvious that the human as a component should be treated with no less respect than the components he operates - the black boxes. It should be obvious also that his performance requirements within the particular system must be considered from the very beginning of the development process.

Field evaluation of a system takes its cue from the development philosophy. For most systems total system effectiveness in field evaluation is an estimate based upon evaluations of components and subsystems of the system. Most field evaluations, therefore, are not, and probably cannot be, evaluations of the system in toto in its intended operational environment. Rather they are evaluations of particular components and subsystems as they operate, in combination with other parts of the system in an environment more or less representative of the intended operational environment of the system. The tested components (both hardware and human) are intended to be representative of those which will finally comprise the total system. These intentions are often only approximately realized.

In field evaluation we have the dual problem of choosing (1) the proper components and/or subsystems for test and (2) the representative environmental conditions under which to test them - this testing being designed to provide an accurate estimate of how well the total system will function in its operating environment.

The fact of total system development being a synthesizing process and evaluation being the testing of components and subsystems justifies laying emphasis on a point made by other writers and one which will be discussed at length in this report. That is, in order to properly evaluate a component of the total system it is necessary that the evaluator know explicitly the role of that component in the total system. The human operator is such a component. The human factors evaluator must know what the operator must do, how well he must do it and under what environmental conditions. This determination cannot be left to last minute speculations by the evaluator in the field.

2.2. The Trouble-Shooting Aspect of Evaluation

A second reason and necessity for evaluation at the subsystem and component level comes about when a particular chain of components or subsystems are tested and the performance fails to meet the standard. Under these circumstances it is necessary to determine which subsystem(s) or component(s) failed to perform to their particular criteria. The human factors evaluator is interested in determining whether the human component failed and, if so, in what way.

These circumstances require that information on performance of subsystems or components be obtained in order to pin-point or diagnose the source of the difficulty. This information must be obtained through a systematic and reliable means suitable for identifying the trouble spot within the larger unit after the larger unit has failed. Thus, there is a trouble shooting aspect to field evaluation which must be made more systematic and reliable with respect to human factors evaluation.

2.2.3 The Adaptive Human Component

A third reason for component evaluation is peculiar only to the human component. It appears to be not unusual during field evaluations that the ingenuity and adaptability of the human operator enables him to perform in a way which results in system success even though his actions and performance may have been quite different from those anticipated by the designer. In such situations his acting in the manner in which the designer had envisioned may not have been within his capabilities and caused him (and sequently the system) to fail. It is this adaptability and ability to recover which characterizes the human component of the

system, makes human factors evaluation (or data collection) important in all systems, and which differentiates the human component evaluation from that of other components of the system. A method for determining when and how the operator has performed in this adaptive way is necessary for guiding re-design, procedural changes or training.

3.0 CRITERIA AND MEASUREMENT

3.1 EVALUATION AGAINST A STANDARD VS. MEASURES OF ABSOLUTE PERFORMANCE

In evaluating a system, component or subsystem standards against which to evaluate are necessary. In evaluating the human component's performance in a given system the test question is not how well he performs in an absolute sense, but rather whether he performs well enough to meet the standards set by the system. However, due to the paucity of human operator performance data, absolute measures of performance should be obtained at every opportunity irrespective and independent of the operator's performance relative to the standards of the system being evaluated. Absolute measures are necessary to provide data on the human component for use in the synthesis of future systems. Collection of this data will continue to be necessary until an adequate bank of human performance design data is established.

This section discusses these two aspects of performance criteria, i.e., that of evaluating with respect to the standard and that of measuring absolute level of performance.

3.2 DERIVATION OF PERFORMANCE STANDARDS

As pointed out in Section 2.0 the synthesis of a system requires the assembly of components which meet certain requirements as to their input-output characteristics and time constraints, and which will operate satisfactorily in the operational environment.

The systems engineer, in synthesizing a system, must be explicit in stating the inputs, outputs, and data transformations required of each component in order for the system to work. Presumably this would also be done for the human component thereby setting standards or criteria for his performance.

However, let us examine the task of the system engineer in designing a system in which he must deal with the human as a component. He is faced with the task of fitting into the system a component about which he has little information as to its exact capabilities for handling data. His guidelines are often general and his data are subjective extrapolations from minimal information. He may know that similar operators have performed similar functions adequately in other systems. He may have available to him the judgments of human engineers as to whether proposed inputs to, outputs from, and transformations by the human component are within its capabilities for handling them. He also has faith in the adaptability (or trainability) of the human component to "come through" and be able to perform adequately. At present the latter may well be his strongest weapon.

The lack of specificity about the human operator's capabilities and the designer's faith in human adaptability often leads to a laxness in stating specifically the inputs to him and the outputs he must deliver. More commonly, it is known explicitly what is required of a chain or combination of components one of which is the human. It has not been customary therefore to enumerate explicitly the data input-output requirements of the operator as a guide and a constraint in designing a system. The result has been specific information about hardware performance requirements and capabilities and the creation of a rather amorphous slot between two pieces of hardware into which the human operator must fit. The required standard of performance for the item of equipment immediately downstream from the operator is known and it is assumed that the operator will perform adequately without being specific as to the level of performance required.

The reader may take exception to the latter statements as over-reaching the facts. Such exception is justifiable to the extent that there are some operator inputs and outputs which can be specified for the system and for which the human capability for handling them can be stated. These tend, however, to be more in the nature of minutiae such as dial legibility, knob size, etc. and concern with them has been characterized as "knobs and dials" human engineering. Attention to these details in design is necessary and their importance is not denied. However, it cannot be denied also that there are a significant number of requirements placed on the human operator by the system and fulfilled by him which are more complex than dial reading and knob turning. It is with these that the human engineer has trouble when he attempts to be explicit about which inputs the operator must and can act upon, in what sequence, what integrative process is required and used, and what the adequacy of the output will be.

As an aside, in this paper concern with knobs and dials, legibility, panel layout and the like will be spoken of as cockpit human engineering and evaluation. Concern with the larger problem of human performance requirements and capabilities as they relate to accomplishment of a mission by a system will be spoken of as mission-system human engineering and evaluation. More will be said of these two later.

Does this amorphism about the human components functions mean that standards cannot be set and performance evaluated? The answer is that his performance can be and is evaluated inferentially when it cannot be evaluated directly. This comes about from the building block nature of the synthesizing process in which components function in sequential fashion and summate into subsystem and system functioning. The validity of inference as to the performance of the human component is a function of the number and variability of the equipment units between the point of his output and the

point of measurement. The standards for operator performance are the standards set for the equipment items into which he makes inputs. This is illustrated in Figure 1 and will be helpful for the human factors evaluator to bear in mind.

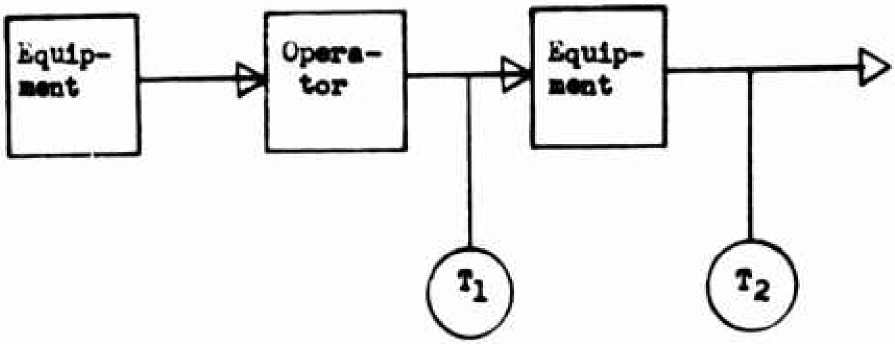


Figure 1. Test points, T_1 and T_2, at which measurements of operator performance may be taken.

In point of fact most human performance data in dynamic closed loop systems is inferential in the sense that the output of some unit of equipment as it functions in response to the operator's input is measured and the human performance inferred from knowledge of the equipments' characteristics. In evaluating human performance in aircraft systems the evaluation is particularly dependent upon inferred performance for many operator tasks. It is important therefore that the human factors evaluator have precise knowledge about the criterion performance required of the chain or combination of elements being tested, of which the human component is a part, and from which he desires to infer the adequacy of operator performance. It is equally important that he know the characteristics of the equipment into which the operator is making inputs and from which measurements are being taken. He must have an assessment of the variability contributed by the equipment to the combined operator - equipment output in order to assess that contributed by the operator.

To recapitulate, two conditions of human operator evaluation are present in the aircraft system. First, that condition in which the performance of the operator can be compared to the required standard (or measured) directly. These are rare for in a strict sense no human performance can occur in the absence of some equipment interface with the human either on the input side, the output side or both. Accuracy in reading information from instruments as

measured by verbal report or time to actuate a given control are examples of test conditions in which direct measurements can be taken. These performances are observed at T_1 in Figure 1. More commonly the human performance is inferred from observing performance of a series or chain of elements as shown at T_2 in Figure 1. Performance is observed at T_2 and inferences as to human performance are dependent upon knowledge of the equipment elements in the chain.

It is apparent that identification, description and priority setting of test points requires a detailed knowledge of the system design. Since many iterations of the design may occur during development, test point description and priorities must be continuously updated. The test points must be made explicit by the contractor's human factors personnel and conveyed to the field evaluation personnel. Further, since field evaluation requires advance planning this information should be conveyed to evaluation personnel on a continuing basis and begin probably not later than the initiation of the preliminary evaluation phase. The specific point during development at which the human factors field evaluator should begin receiving this information cannot be pin-pointed exactly. It will vary with the system, its uniqueness, its complexity, etc. However, it is certain that it cannot be left to be determined by the field evaluator alone just prior to reception of the system by the test facility as is the case with most systems at present.

The specific test points and the required levels of performance must be identified through the system analysis and synthesis carried on during development. As the system design becomes more firm and the hardware characteristics become fixed, operator-equipment interfaces must be identified and those output points selected from which operator performance can be inferred through measurement of equipment outputs.

3.3 ABSOLUTE LEVEL OF PERFORMANCE MEASUREMENT

Performance data obtained during field evaluations must serve both the need of evaluating a particular system and that of providing data useful in the design of future systems. Therefore, it is important, both for the evaluator in assessing his data and for future use of the information, that the characteristics of the equipment with which the operator interfaces during the test be known and reported. It is one of the weaknesses in the data on human performance that there is incomplete knowledge about the specific characteristics of machines and equipment with which the data were collected and which correlate with performance. In most cases, the human engineer cannot be as exact about what parameters of a piece of equipment will affect operator performance as, for example, the aeronautical engineers can be about what parameters will affect the stability of his aircraft. It is therefore incumbent upon

the human engineer during design and the human factors evaluator during system test to enumerate and describe the parameters of the equipment and conditions of test if the human performance information obtained is to be useful for future design.

3.4 AN EVALUATION MISSION

In planning the field evaluation and delineating test points, standards of performance at these points must be set. However, these standards may differ depending upon the overall mission being undertaken and by segment of a mission. It is necessary then to formally define the evaluation mission in order to arrive at test point standards.

In setting up an evaluation mission it is suggested that it be built around the primary mission of the system with emphasis upon the weapons delivery phase. While test points should be identified during the more common phases of the mission such as launch, climb out, cruise, loiter, etc., for the types of aircraft with which we are concerned, the primary emphasis and priority should be upon that phase of the mission in which the enemy is encountered or the target attacked. Special problems may arise, of course, with less conventional aircraft which demand special priority but even there the primary weapons delivery role of the system should be emphasized.

In selecting the evaluation mission a suggested further consideration is that "worse case" conditions be represented within it. By worse case conditions is meant those conditions in which the greatest demands are placed upon the human operator either as to level of performance or time pressure. Further, as a guide to the selection of the evaluation mission consideration should be given to the type of mission which will be flown in the systems' most likely theater of operation.

The evaluation mission must then be broken down into segments. The functions carried out during each segment and the equipment involved must be identified from the systems and task analyses carried out during development. Those functions in which the human component is involved are then specified. Decisions must then be reached as to the relative importance of the functions to mission accomplishment and the test points identified for obtaining performance which will give the most valid picture (either directly or inferred) of the operator's performance.

3.5 ASSIGNING PRIORITIES TO TEST POINTS

Identification and description of test-points are critical to human factors evaluation. However, since it is not practically possible for measurements to be taken at every identifiable point a priority must be set up. Two bases for priority are necessary.

One, a priority system is needed which ranks the relative importance of the test point performances with respect to their effect upon overall mission success. Second, a system which ranks the test points relative to the degree to which assumptions have been made about human performance during design is needed. Those cases in which the least data were available on which to predict human performance when he was designed into a given slot should be given highest priority in testing in order to test the assumptions. Parenthetically, these points in particular should be tested as early in the development process as possible.

The relative importance of the several functions to mission accomplishment may be obtained by the method described by Rook (1964) using rank order techniques. For guidance in the use of rank order and paired comparison methods the evaluator should have at hand a copy of Guilford, 1954. Rook describes his procedure as that of describing concisely on separate cards each of the events in the group to be evaluated. For paired comparison the cards can be presented in pairs and the judge asked to make a judgment as to which is the most and which is the least important to mission success. If the number of events to be ordered is too large for the paired comparison method to be feasible the cards may be simply ranked from most to least important.

A method for ranking such events which the present writer has found feasible for use in the field and which captures the judges interest combines the paired comparison and rank order methods. It is recommended particularly when the number of events to be judged is large. With this method the judge selects an event (a card) which he feels lies at some point near the middle of the range of events to be judged. He then compares each card with this reference card and sorts the cards into two decks one containing those events judged to be more important than the reference event and one containing these judged to be less important. Each of the resulting decks are then sorted in the same way. This sorting procedure is carried on until each deck is small enough for the events within the deck to be conveniently ranked within the deck, i.e., 2 to 5 cards. The result is a ranking from most to least important of all events within the total deck. It will be noted that this latter technique allows the judge to make a comparative judgment and forces transitivity on the scale resulting from judgments of the attribute. It then avoids the possibility, as is the case in paired comparison, in which A may be judged greater than B, B greater than C, and C greater than A. The serious student of scaling methods should consult Torgerson (1960) as well as Guilford (1954).

No ranking or ordering procedure can result in inter-judge reliability without a clearly understood statement of the underlying attribute being judged. Thus the importance of defining what is

meant by "importance to mission success" cannot be over emphasized. Rating methods and scaling techniques cannot be discussed at length here but the necessity for their being taken up in indoctrination courses for human factors field evaluators is emphasized in the course outline given in Appendix B.

3.6 HUMAN ENGINEERING PRINCIPLES AND HANDBOOK DATA AS CRITERIA

The criteria most used during design and evaluation are human factors design principles and data contained in various Human Engineering Handbooks. The applications of these principles and data throughout the development period and throughout evaluation serves a useful purpose. Their usefulness decreases, however, as the system moves toward completion and dynamic full system evaluation. The blind application of these principles and data without careful consideration of the total system configuration, the systems mission and operating environment can result in incomplete, if not, misleading evaluative conclusions.

As a general rule the human factors design engineer and evaluator should be guided by the principle that the system is being designed to optimize its capability for executing specified missions. This rarely allows for optimizing the operator station ideally to fit human capabilities and provide for his comfort. Space and weight are at a premium particularly in aircraft and compromises are the rule in fitting components together in order to bring about total system-mission functioning. These compromises are made toward the end of configuring the system to accomplish its end goal or purpose. As a consequence, the evaluation of any component within a real system cannot be in terms of some ideal conditions under which it would operate best. Rather it must be in terms of the best conditions possible within the real constraints laid on the system. Thus human engineering principles used as evaluative criteria by way of a check-list or similar device has limited utility per se.

They have utility only when used within the context of real system design constraints and system purpose. Similarly, the extrapolations of handbook data must be made within the same context. The use of either in the absence of knowledge of the overall system contraints and purpose can lead to quite erroneous conclusions as to the adequacy of the system design.

Human engineering principles and handbook data are generally used as design and evaluation criteria early in system development. They can be used most appropriately by those human factors engineers working closely with equipment design engineers and with those most familiar with the system mission and its constraints. It is very difficult for anyone not completely familiar with the details of the mission, component compromises and system constraints to apply successfully a human engineering check list to the evaluation of an operator station.

The successful use of these criteria early in development to evaluate the operator station and thus, by extrapolation, to predict how successfully the operator will perform in the system is termed "cockpit oriented" evaluation in this report. Here the operator station is examined with respect to work-place layout, design of controls, functional placement of displays, control-display relationships, illumination, anthropometric measurements, visibility and the like. Assessments are then made as to the adequacy of the design in meeting these "compatibility" criteria. Such an evaluation is usually first carried out on the cockpit mock-up and repeated, although less formally, on into fleet use. Unfortunately at this time, this type of evaluation may be essentially the only type of human factors evaluation carried out for a system. It is the only type that can reasonably be carried out, either formally or informally, when the evaluator does not have detailed information as to the mission of the system or the test points at which empirical performance data may be obtained.

In one of the systems examined by this writer the yellow sheets turned in by the "field" evaluators of the system contained over 85% cockpit oriented items. That is to say that the evaluators were concerning themselves with such items as parallax in reading an instrument, inability to reach a control with harness locked and the like with no concern reflected in the yellow sheets for operator-system capability for executing missions or mission segments. Of course, one explanation for such a lack might be that the operator-system combination did function to mission standards. A more likely explanation is that the evaluator was not given information as to the mission which would allow him to evaluate it against mission criteria - which happened to be the case in this instance. Without such information and without an orientation which would lead him to view the human operator as a component which must deliver up certain inputs to a given level of accuracy in a given time the evaluator understandably carries on in the tradition of the mock-up inspection during his field evaluation.

It is important that all concerned with systems evaluation come to the realization that human factors evaluation is more than the application of "good human engineering practice" to the cockpit. They must begin thinking of the human as a functioning component whose capabilities for procession information, as he is designed into the system to process it, must be evaluated. Cockpit oriented types of evaluation are useful early in development if carried out by the proper people and with proper concern for equipment constraints and system purpose as outlined earlier. Given these considerations it is difficult to see how the mock-up inspection procedure as it is presently exercised could possibly produce an evaluation with any substantial validity.

3.7 MEASUREMENTS AND CRITERIA

Two aspects of performance relate to mission success and are important in human factors evaluation. Mission success will be compromised if a component fails to perform either to the required accuracy or within the required time.

3.7.1 Accuracy Measurements

The appropriate metric for use in assessing human component performance when he functions within a chain of equipment components must be arrived at by examining the outputs from the equipment at the designated test-points. Human component performance is inferred from measurement of the equipment output to which the human component must provide the input. This will be the usual case in field evaluation and often those measures of interest to the equipment engineer (especially during Board of Inspection and Survey (BIS) Trials) will be appropriate for assessing human performance. This is not to imply that the measurement values obtained by the equipment engineer during his tests will necessarily be useful to the HF evaluator. The equipment engineer is often interested in equipment performance in response to standard or stylized inputs which may not be fully representative of the way in which the equipment will be used by the human operator in carrying out the evaluation mission. Nevertheless, the measurement parameters and measurement tools used by the equipment engineer can prove highly useful to the human factors evaluator.

The responsibility of the HF evaluator is to select those measurement parameters from the engineer's arsenal which most directly reflect the human component performances required at the test-points identified earlier and to apply them while the system performs the evaluation mission. The human factors evaluator will need to pay particular attention to those parameters which test those assumptions about human performance made during systems design which were made on the basis of little or no data or required rather extensive extrapolations from available data.

An hypothetical example using equipment which nudges the limits of the state-of-the-art may serve to make the responsibilities of the human factors evaluator more clear. Suppose that a direct lift control system were designed into an attack aircraft for use during carrier approach and landing. The equipment engineer, in testing the system will be interested in determining whether the output, measured in altitude change in response to prescribed standard inputs, is as he predicted when he synthesized the system. However, since he synthesized the system with a human component in the chain he made certain assumptions about the characteristics of the human component for receiving inputs and making outputs within the given time constraints. The human factors evaluator, using the same measurement parameters and tools can test these assumptions under conditions representative of the operational mission, i.e., actual carrier landings. More often than not the human factors

and the equipment engineer can work together in combining their tests to mutual benefit and improvement. This is particularly applicable during the earlier stages of development.

It is particularly important that the teaming up of human factors and equipment engineer during design and evaluation be appreciated. It is important to remember also that whenever it is possible, a third member of the team may be very useful. This third member is an experienced operator of currently operational systems which are similar to the system under test. This member brings to the evaluation a knowledge of the operational requirements and constraints to which the system will be subject. Again it should be emphasized that it is important that he be assigned to the team as early in the development sequence as possible.

What can be said about specific measurement parameters in human factors evaluation? It should not be necessary to warn against selecting those measures which can be conveniently obtained or nicely quantified unless they also meet the criteria of importance to mission success. At the same time, the Human Factors evaluator should not be reticent about insisting on particular measures or special equipment in order to assess the human component. Every component in the system must be evaluated at one time or another during the development of the system. The human component deserves no less consideration than any other.

As to specific measures these will vary by system, at different stages of system development and with the evaluative tools available. With the system of concern here, i.e., the fighter and/or attack aircraft, a frame of reference is suggested which will help to insure inclusion of the appropriate measures. In this frame of reference three major areas of concern are broken-out. These are (1) aircraft attitude control, or stabilization about the aircraft axes, (2) navigation, or position and translation in three dimensional space and (3) system state, or the condiction of power plant, armament subsystem etc.

Within the first category it should be first kept in mind that the system is designed to be a platform or carrier for various ordnance. As such it must attain certain attitudes and/or be stabilized for proper delivery of the weapon. It will have been assumed during design that the human component can provide inputs into the system which will control the attitude of the aircraft within limits which will result in weapon release with high probability of hitting the target. This assumption may have been only implicit but, nevertheless, was made and must hold for mission success. Thus, angular deflections and rates in pitch, roll and yaw will be required to be maintained if successful weapon delivery is to be attained. Therefore an assessment of the human operator's performance in maintaining them within the required limits is necessary.

In some instances the human factors evaluators' task could be that of synthesizing data already obtained on the system from other sources. For example, if the required limits within which aircraft body axes deflections and rates must be maintained for weapon delivery were stated and if data were available from handling quality studies which provided quantitative data on operator performance, an assessment could be made of weapon delivery success with the human operator in the loop. Here the team effort mentioned earlier is essential. Unfortunately for this example the handling quality data is obtained using highly experienced and well trained operators. Their performance data, therefore, could not well be considered representative of that of the fleet pilot without substantiating information.

With respect to the second category, navigation, a broader meaning is intended than that ordinarily ascribed to the term. Here the concern is with control of the position and/or rate of change of the vehicle in three dimensional space. Therefore, control with respect to these space axes during carrier approach has the same dimensions of measurement as during cruise. The standards or limits within which control must be exercised, however, will differ by mission segment.

The third category, system state, may be thought of as the condition of the system at any moment in time apart from its attitude and spatial position. In this category are the requirements that the power plant output, armament system set-up, etc. be in a given "state" at given times. It contains both monitoring (perceptual) activities and procedural or set-up activities of the operator. Of particular concern in current attack aircraft is the problem of the armament system being set-up in the proper state for given conditions of operation.

In brief, certain measures are common to each of the above categories with the standards required varying with particular segments of the evaluation mission. Parenthetically, in some instances it can be inferred from satisfactory operator performance on a parameter during a given mission segment that performance on that parameter will be satisfactory during another segment. For example, from satisfactory performance under worst case conditions satisfactory performance under less demanding conditions (in terms of standards by accuracy) may be inferred.

The measurement parameters appropriate to each of the categories are discussed here in a general way to provide an orientation for the evaluation for aircraft systems in general. A matrix of specific measures suggested for use with the attack aircraft is given in Appendix A.

For the first category (attitude control and stabilization) measures of the pitch, roll and yaw angles and rates of the aircraft may be used to infer the operator's ability to maintain these parameters within required tolerances. For modern fighter and attack aircraft, engineering knowledge in the areas of control theory and stabilization systems plus heavy reliance upon the prior training and experience of the transitioning pilot has resulted in few apparent problems in this area for the evaluator. This situation may be more apparent than real.

The test methodology for determining whether the aircraft meets certain performance specifications and exhibits certain handling qualities seems fairly well established. At least in the hands of experienced and specially trained test pilots the performance limits and handling qualities can be relatively well established. The methodology for determining whether the fleet pilot can handle the aircraft within limits while setting up his armament panel, executing evasive tactics and performing various and sundry other chores is not established. Obtaining the judgments of experienced test pilots in evaluating the basic performance and handling qualities is a necessary and important first evaluative step. However, this step must be considered as a test of the performance of the hardware rather than an evaluation of the performance of the human component for which the system was ultimately designed.

Therefore, measurements of aircraft attitude parameters, when the aircraft is flown by fleet pilots under conditions at least approximating the weapons delivery segment of the mission, are necessary to obtain evaluative data on the human component of interest.

For the second category (navigation) the argument just presented holds also. When the experienced test pilot tests the system, that test must be considered essentially a test of the hardware. As a test pilot he tries to control and standardize his inputs so as to determine the operation of the equipment he is controlling. Once the adequacy of the equipment is established it remains to be determined whether the fleet pilot working with that equipment, can produce the requisite performance to bring about successful system performance. The latter is the concern of the human factors evaluator.

In the navigation category, evaluative measures take as their standard and starting point the notion that the aircraft, at any given moment in time, must be at a given position in three dimensional space and/or the first and second derivatives of position must be within certain tolerances. That is the operator must function to control the position of the aircraft, its rates and accelerations to be within specified limits. These limits must be determined for each segment of the mission, e.g., carrier approach,

touchdown, etc. During development these criteria will necessarily have been stated or assumed in synthesizing the system. Their explicit identification and the establishment of test-point priority have been discussed earlier.

The third category, system state, including as it does monitoring and procedural activities, is highly important in the present generation of fighter and attack aircraft. The diversity of ordnance capable of being carried places a premium on the operator's ability to carry out armament panel set-up procedures under severe time constraints. It is in this category that a number of decisions and procedural activities are likely to be relegated to the pilot as development progresses and new capabilities and tactics are foreseen for the aircraft as a weapons system. The pilot's ability to carry out the functions in this area are more generally indicative of his ability to use the aircraft in its intended role as a weapon while the first two categories apply more generally to his ability to control it as an aircraft. The human factor evaluator will need to be concerned with this category and test points within it will tend to have high priority for mission success.

3.7.2 Time as a Measure

In the course of this study it has become apparent (at least to this investigator) that an important and usually critical parameter for evaluation is that of time. In the systems of concern in this study conditions change rapidly and information update rate is often critical to successful system performance. In actuality this is true of many systems other than fighter and attack aircraft. Therefore, a critical question to be asked in the evaluation is whether or not the human component can execute the required operations in the prescribed time.

The equipment engineer is likely not to regard time as a critical factor in the sense that he worries about his components reacting too slowly. However, it is important to recognize when man is introduced into the system that he does not react with the speed of equipment components and often may take a critically long time to perform a given function. In fact, our major concern with man in today's complex system may be more with his inability to perform all of the functions in time rather than his inability to perform accurately.

In singling out time as a measure of performance it is assumed that the human operator, unlike the black box components, recognizes the need for accurate performance in order to insure his own safety and the success of the system (or both) and strives to achieve it. Thus, time to execute becomes an important standard against which the human component must be evaluated.

The time limits within which the total system must achieve a mission or mission segment can be relatively objectively fixed. The number and sequence of operations required of the system components within this time frame can also be enumerated. Again, whether the human component executes his required actions within the time limit may be measured directly or inferred depending upon the location of the test point within the component chain. As with accuracy measures, the conditions of test and the equipment characteristics must be completely described in order that the data will be maximally useful for future designs. Also, whenever possible, actual time to execute measurements should be taken irrespective and independent of whether the operator performed the function within the required time, again with the object of obtaining data useful to future system design.

4.0 LEVELS OF EVALUATION

4.1 THE ROLE AND DEFINITION OF FIELD EVALUATION

In any study of the human factors evaluation problem, restricting the study to "field" evaluation creates a certain dilemma for the investigator. First of all he faces the difficulty of defining where field evaluation begins and where it ends. For although there are formally defined stages in evaluation, e.g., BIS trials, in actuality evaluation does and must being much earlier and extends into operational use of the system.

Secondly, there must be some definition of what field evaluations are to do. Broadly speaking they are meant to test the system against assumptions made about its performance when it was originally conceived. In a very real sense any test which predicts how well the system will perform is desirable at whatever point in the development process it is conducted. The element which is added to test and evaluation through conduct in the field is presumably that variables and conditions are more nearly representative of those in fleet operations and therefore more valid. This representativeness varies from system to system depending upon similarity of the system to previous systems, urgency of system need, availability of test personnel and equipment and the like.

Field tests conducted late in or at the end of the development period generally will have greater "content" validity than those conducted earlier. That is to say that the equipment to be tested and the conditions of test will be judged by competent evaluators to be good likenesses of the ultimate criteria - fleet use in operations. However, deficiencies found in the system at this stage are more costly and difficult to correct. It is a common feeling among evaluators that by the time the system reaches the field evaluation point the system design is "set in concrete" and quite resistive to change.

This inflexibility is particularly injurious to the human factors component since the human operator is viewed in a different light from hardware components. Being adaptable it is often felt that he can "learn to live with" a deficiency or it can be corrected through training or special selection of the operator. A change in the system at the field stage to better accommodate the human component has been most difficult unless it is declared an item affecting the safety of the operator. It is desirable, therefore, that valid evaluation of the human component performance be carried out as early in the development process as possible. (Parenthetically, this is true for any component). The more valid the component testing early in development the fewer problems that will arise during field testing or subsequently in fleet use. The most desirable situation would be one in which tests carried out at various stages of development have high predictive validity for predicting performance in the fleet.

This predictive validity in which empirical relationships between tests and fleet performance are established cannot be obtained until reliable measures of performance in the fleet are possible - something toward which work should be directed but not within the purview of this study.

However, at some point before actual fleet deployment an evaluation must be made to determine whether the system performs to the original conception and specifications. The results of this evaluation can and must serve as criteria against which to validate tests conducted earlier in the development process. Validated tests conducted earlier in the development would reduce human factors system deficiencies at this accept-reject point and deficiencies could be corrected at a time when the system design is more amenable to change.

4.2 EQUIPMENT EVALUATION VS. EQUIPMENT-OPERATOR INTERFACE EVALUATION

The system equipment undergoes almost constant evaluation throughout development continuing up through Technical Evaluation. In many ways the equipment designer can test his equipment under conditions which equal or go beyond the constraints and conditions of its operational use. During these evaluations variables can be and are rather carefully controlled. When a human operator is used in these equipment evaluations he is trained and instructed to make carefully controlled inputs to the system and to make prescribed observations. Emphasis is on equipment functioning and test which is reasonable and necessary. It should be recognized and remembered, however, that these are evaluations of equipment and not of operator-equipment interfaces as they would occur in fleet operations using fleet operators. The human operator's performance as he interacts with the equipment is not being evaluated through measurement at specified equipment test-points. (These test points have been discussed in Section 2.2). Rather the equipment is being evaluated as it functions with an operator closing the control loop but an operator with as standard inputs as possible in order to get a clear picture of equipment functioning. These are Technical Evaluations and have been considered a necessary prerequisite to operational evaluation in which the system is exercised more nearly as it will be in operations.

From these equipment oriented evaluations some information relative to the human component is obtained. However, the information tends to be cockpit evaluation oriented rather than mission-system oriented. The experienced test pilots who work with the system up through Technical Evaluation can and do make observations as to cockpit design deficiencies and yellow sheets are generated. However, it is difficult to get the test pilot to think "mission" at this stage when his job is really to test the assumptions and

specifications relative to the equipment. Theoretically, at this point the human factors aspects must also be evaluated but it is most difficult for the test pilot to put himself into the place of the operational pilot and to make observations relative to the adequacy of the operator-equipment interfaces in carrying out mission functions. He can and does make observations as to such things as control accessability, lack of visibility, difficult in carrying out procedures and the like for the mission segments which he performs. That is to say that, in his Technical Evaluation, he will perform the mission segments common to fighter and attack aircraft such as launch, climb out, cruise, approach, etc. During these evaluations he is in a position to observe deficiencies both in the equipment functioning and in cockpit design which make his job as an operator difficult or less than optimum. However, the man-machine adequacy in performing these mission segments is assessed only with a highly experienced operator in the loop.

4.2.1 Test Point Measures

During Operational Evaluations human factors evaluation takes on more of the mission oriented and less of the cockpit oriented flavor. The components and subsystems (component chains) can be tested for their adequacy in performing functions which relate more directly to ultimate mission purpose, i.e, delivery of ordnance. The system synthesis process and the necessity for component and subsystem evaluation has been discussed earlier in Section 2.1. In testing the human component, the use of test points downstream in the component chain from the human component has also been discussed. Further it has been suggested that these test points in the total system be identified and their relative importance to mission success be established so that during total system test, either direct or inferential human performance data can be obtained. At these test points standards of performance must be set within which each component must operate in order for the combinated component performances to summate into total system success. These standards will vary with the purpose (mission) to which the system is put and by mission segment. The definition of an evaluation mission has also been recommended in order that conditions of test may be standardized and repeated measures taken under uniform conditions. The suggested parameters for obtaining data at these test points are given in Appendix A.

4.2.2 Trouble Shooting

When subsystems or chains of components are tested and they do not come up to standard it is necessary to determine where the problem lies. When measuring at some test point at the end of a chain of components in order to infer the adequacy of the human performance, we assume that the equipment is functioning properly and that any inadequacies discovered lie in the human output.

However, the focus of the malfunction within the component chain is not always clear and cannot be reliably attributed to human component failure. At times the test point may be at the end of a complex combination of components such that it is not easily inferred as to where the failure lies. In such cases it is necessary for the operator to be trained in and have the necessary tools for reliable retrospection or concurrent observation of human component problems as they occur. These observations serve to trouble shoot the subsystem and point to the deficiency in the design which leads to inadequate performance at the operator-equipment interface.

At present this type of information is not obtained in any systematic or formal manner. Some test pilots record their observations incidental to their report of their observations of equipment functioning. It has been reported to this investigator that often an operator-equipment interface problem is not reported since the pilot feels it to be something that would happen only infrequently and probably not at all after the operators become familiar with the system. At other times they are not reported because the test pilot feels that a problem which has arisen reflects upon his own ability as a test pilot or is not typical of his performance so he tends to overlook reporting it.

A formal list or guide is needed to systematize the pilot's reports with respect to deficiencies which are not possible of being detected from measurements of the subsystem performance in order that these problems can become a matter of record and a part of the formal evaluation. These reports should be accompanied by a complete description of the conditions under which the deficiency occurred, the equipment from which the operator received his inputs and to which he made outputs and his level of experience and familiarity with the system. This type of data can then be accumulated as a data bank for guidance in the design of future systems as well as guide correction of the deficiency.

4.2.3 Operator Adaptation

The test pilots must be especially indoctrinated to the problem of evaluation which arises when the subsystem being evaluated functions properly but the human operator has had to perform in an adaptive way or to recover from a mistake in order to make it operate properly. If measurements are being taken at a test point and the subsystem operates to standard there is no way of detecting performances in which the pilot adapted or recovered from errors and thus caused the subsystem to function correctly. Therefore a formal and systematic method is necessary such that the test pilot will report in detail such behavior along with the conditions of the equipment and his level of experience and familiarity with the system. These reports may lead either to a redesign of the system in order that such adaptive behavior is not necessary or may be incorporated into training programs or procedural changes.

Indoctrination and training of test personnel is discussed in Section 5.5 and a suggested course outline is given in Appendix B.

4.3 VALIDATION AND USE OF EARLIER TESTING

It has been pointed out earlier that discovering man-machine interface deficiencies at the Techeval or Opeval level makes them difficult to remedy since the system is so far along in the development process. It has also been stated that the Techeval and Opeval must serve as a substitute for the ultimate criteria of fleet operations (at least at the moment) as well as serving the function of testing the assumptions made during the conception and development process. In this intermediate criteria role the Techeval and Opeval can serve as criteria against which to validate testing carried out earlier in the development process. The earlier valid testing can be carried out the greater the opportunity to influence design. These earlier tests take a variety of forms.

4.3.1 Mathematical Models

The high speed digital computer has opened the possibility of using mathematical models to manipulate system variables and determine the effect upon total system operation. In their use mathematical models are set up to simulate parameters of the system so that values of the parameters can be varied and interactions among parameters can operate.

These models can be categorized as falling into the classes of (1) "time available" models, (2) reliability or probability of failure models, and (3) error and importance of error models. Their limiting feature at the present is the paucity of relevant data. Their appeal lies in their ability to handle a large number of parameters and their interactions and to very quickly determine how changes in these parameters will affect the system operation. Their potential contribution to testing early in development is great, however, at the moment no impressive data exists as to their validity.

No recommendations are made in this report with respect to any particular model. The contribution of the present work to the model approach can be greatest in setting up of methods and techniques for establishing test points and data gathering and insisting that the conditions of test be fully described so as to make the data maximally useful in modeling future systems.

4.3.2 Physical Models

Physical models of the ultimate system range from simple static mock-ups of a part of the system through dynamic simulators to the aircraft itself. This section discusses these models and their role in human factors evaluation.

4.3.2.1 Static Mock-Up - At present the physical model most used in human factors design and evaluation is the cockpit mock-up. Since the mock-up is both well used and mis-used it deserves special discussion.

The mock-up is used intensively and almost continuously by design personnel as an evaluative tool during the design process. In this role it is both useful and necessary. However, the static mock-up is the focal point for formal human factors evaluation which, if not outmoded, needs improvement. The successful evolution and validation of mathematical models implemented on the high speed computer may eventually replace and go much beyond many of the functions now fulfilled by the static mock-up. In the meantime, it is believed by this investigator that major improvements can be made in the human factors evaluation procedures and techniques which are used during mock-up inspections as they are now constituted. This belief has been found to be almost universally supported by those having experience with the mock-up inspection.

As a background for discussing the mock-up inspection and the role it plays in overall evaluation it is necessary to discuss two levels of evaluation. The first level is that which was indicated in an earlier section as being "cockpit oriented" evaluation. The second was referred to as "mission-oriented" evaluation.

At the first level of evaluation the cockpit mock-up is examined for what may be termed compatibility with the operator's capabilities and limitations. The operator station is examined in the light of human engineering design criteria such as conformity with good human engineering principles and with handbook data. This means evaluating on the basis of work-place layout, control coding, control-display relationships, illumination, anthropometric compatibility and the like. Assessments are made as to the adequacy of the design for meeting these design principles and handbook data criteria. The conscientious evaluator will also evaluate on the basis of functional arrangement of displays and controls. This type of evaluation is cockpit oriented evaluation.

To carry out a human factors evaluation of this type there are certain requirements with respect to personnel, method and techniques. Personnel must be knowledgeable with respect to the human factors literature and data, be able to critically evaluate that data, and extrapolate from it in the light of the particular system being evaluated. Further, they must be cognizant of the principles of good human factors design insofar as these are known and can be applied to the specific system being evaluated. The areas and points to be covered during the evaluation should be incorporated into a checklist such that a systematic examination of the operator station may be made and insure that important areas are not overlooked.

It is particularly important that the human factors evaluator be knowledgeable with respect to the personal equipment to be worn by the operator and that evaluations, particularly with respect to work space layout and anthropometrics, be conducted taking into account the effect of personal equipment upon performance. This latter point seems minor but nevertheless is sometimes overlooked and can result in major problems.

No disparagement of this type of evaluation is intended if carried out properly by competent and informed personnel. Indeed, it is a necessary antecedent to evaluations of the second type.

The second level or mission oriented evaluation requires an examination of the operator station in the light of operator functions and tasks and an assessment of whether the system will do what it was designed to do to some specified criterion level. In order to carry out such an evaluation certain data and information must be known over and above that required for carrying out a cockpit oriented evaluation. It is necessary to know in detail what functions and tasks the operator is required to perform, in what sequence, and to what criteria. These criteria are phrased in terms of performance standards, sequence of operation and time. That is to say that the determination must be made as to whether the operator can carry out his functions and tasks in the proper sequence to the required accuracy within the required time.

Any mission oriented type of evaluation requires the generation of very specific and detailed data regarding the system. To carry out such an evaluation the evaluator must have specific and detailed information about subsystem functions, data flow, and the requirements placed upon the human operator with respect to sequencies, accuracies and time for carrying out his operations. He must have this information as it applies to the various segments of the system mission and certainly for the evaluation mission if one has been generated.

The static mock-up as a physical model serves very well as a tool in helping to make decisions and evaluations about physical arrangements of the cockpit. It is not an adequate tool for the mission oriented type of evaluation in which assessments are made of whether or not the operator will be able to perform to the standards required by the mission - a more critical requirement.

It is relevant to raise the question as to the advisability of a mock-up inspection in which the static mock-up is evaluated and a binding decision made as to its configuration. The advisability is questioned because of difficulties with it experienced by this investigator and echoed almost universally by others with mock-up inspection experience. The first difficulty is that dynamic man-machine interactive performance cannot be evaluated. This, of course, the static mock-up evaluation has made no pretense at doing.

However, this performance relative to the required standard is the central question particularly when it is recalled that the operator station must be configured to optimize total system functioning for performing a mission rather than optimized to conform to the operators capabilities and limitations.

Before the advent of such complex systems as we now have with their wide variety and sophistication of armament the mock-up inspection was a natural procedure and served a useful purpose. In simpler systems whose effectiveness and employment were much more dependent upon the ingenuity of the operator the cockpit layout for operator convenience was undoubtedly of major importance. In the complex systems of today the operator is required to perform more a dynamic "component in the system" role and functions to process information in a standard way toward the purpose of bringing about a specific end product through the employment of specially designed equipment. The static mock-up does not lend itself to such an evaluation and in a real sense is vestigial having not only outlived its usefulness but prevents, by the inertia of its use, the adoption of more effective procedures.

The second difficulty is that in order to be useful cockpit oriented evaluations, the mock-up inspection team personnel must have the qualifications outlined earlier. That is, they must be cognizant of the principles of good human engineering insofar as these are known and can be applied to the particular system being evaluated. This is too often not the case.

Mock-up inspections have been carried out by teams composed of personnel representing the various specialities concerned with the system. Judgments and evaluations as to the adequacy of the human factors design generally are made by any member of the team. Often these judgments reflect the particular experience and preferences of the individual rather than the considered application of known principles.

A third difficulty is that in order to even approximate an evaluation which is mission oriented during a static mock-up inspection the team member must have important additional information. In addition to knowledge of general human factors principles and a checklist to determine systematically the adequacy of design from the cockpit orientation point of view, he must have detailed information with respect to specific tasks to be carried out by the operator as he functions as a component within the system in performing the mission. He must have this information as it applies both to normal and "worst case" operation of the system preferably within the context of a well detailed evaluation mission. He must also have intimate knowledge of the equipment with which the operator interacts to include detailed information on data flow, flow rates and required outputs from the operator. He must then project himself into the role of the operator functioning to handle these data to bring about mission accomplishment - a procedure not likely to produce high predictive validity.

It is virtually impossible for an inspection team member to have such familiarity with the system when he comes to the mock-up inspection. Nor is it possible for him to acquire a sufficient appreciation of it during the inspection time period. The net result is that the inspection team members make human factor evaluative judgments based upon limited knowledge of the intricacies of the equipment and details of the mission. Their judgments therefore tend to stem from their own personal likes and experiences. They must rely upon the orientation and guidance from contractor human factors personnel who are familiar with the system from their work with it and upon available human factors analyses data. The short period of a mock-up inspection is not sufficient to gain a comprehensive picture of the system and its requirements.

A third alternative is recommended which would prove feasible with most systems. Since great familiarity with the system equipment and mission is required the assignment of customer personnel to the contractors facility during development is suggested. These personnel would be experienced operators of the prior system and would provide expertise in operational problems and usage of the system. In turn they would learn the details of the equipment with which the operator must work and what the system was designed to do. Working together with contractor personnel the cockpit design would be evolved to be "frozen" at a designated point during the development process. These customer personnel would represent the human factors area during the mock-up inspection and be members of the inspection team. The requisite qualifications of these personnel are discussed in greater detail in Section 5.3.

4.3.2.2 Simulators - The simulator as a dynamic physical model of the cockpit holds many advantages over the static mock-up. From the evaluators' point of view it offers flexibility, opportunity to obtain reliable performance data and a test situation more representative of the real system.

It offers an opportunity to provide data to the later "field" evaluations which may lighten their load and even go beyond what is possible in the aircraft itself. The use of such data is encouraged by Board of Inspection and Survey Aircraft Test Directive No. 1 - 6, 17 September 1965 as follows.

> 3. Policy. It is the policy of the Board to accept for trials purposes data from any source provided that in the judgment of the activity conducting the trials the data are valid and are fully representative of the production article undergoing trials.

It has been reported to this investigator that the simulator cannot be built in time to be useful for most systems. This disadvantage appears to be lessening as designers become more knowledgeable about simulator requirements and more of them come into existence. More and more they are coming to be used to solve special design problems and to evaluate decisions about particular subsystems.

The Heads up Display (HUD) is a case in point. It is also pointed out that simulation is much more expensive. However, it is also acknowledged that it takes only a few ECP's or aircraft accidents to pay for the simulator cost.

An additional advantage of using a simulator during design is the knowledge to be gained about training problems, training procedures and training device configuration. It is difficult to estimate the time and money that this saves. Parenthetically, the experienced customer personnel assigned to the contractors facility (recommended in Section 4.3.2.1) can also contribute greatly to training plans and training equipment specifications if they are given the opportunity to participate in design and evaluation using a simulator.

The simulator is demonstrating its effectiveness more and more during design and evaluation. Its use during mock-up inspection as those inspections are now conducted would give the inspection team members greater insight into the human factors design. However, its effectiveness would be much greater if used on a continuing basis by contractor and customer personnel to arrive at a suitable configuration. The simulator to be most effectively used requires personnel trained in the conduct of evaluation using such tools. The customer personnel could contribute greatly to identifying the important parameters, conditions and constraints to be considered during any test.

Simulator tests also requires the collection of data over some period of time using representative operators. The data requires reduction and analysis which, while becoming less and less of a burden with modern computer technology, nevertheless does not lend itself to the "spot check" type of evaluation characteristic of a mock-up inspection.

4.3.3 Part Task Testing

The initiation of human factors evaluation from the beginning of the system development with the assignment of manufacturer and customer personnel to conduct and/or monitor these evaluations allows for the introduction of other evaluative techniques. These techniques are those which we have termed "open loop" tests of display design configuration and which may be used in weeding out or narrowing down design alternatives in early stages of development. These techniques are essentially tachistoscopic presentations of display designs used for comparative evaluation. They are, in fact, the techniques which were used largely in the laboratory setting, in obtaining a major portion of the human factors data available to us today.

The advantage in using such techniques as a part of the development of a system over its use in the laboratory is that it would be used by those familiar with the problems, variables and constraints peculiar to the system of interest and would generate data applicable to system needs. The technique has been tried in a limited way in connection with system development. It's point of entry into the evaluation process is shown in Figure 2.

The open-loop testing technique is termed "open-loop" since the response of the subject has no direct effect upon the next stimulus presentation. The stimulus material may be a display configuration presenting information which the subject is required to interpret, readout, and report. Measures of the subjects performance may be the speed and accuracy with which the information is readout. The stimulus material may be used in the form of a static projection, i.e., a 35 mm slide, in such case it is termed static open-loop testing. The material may be presented through use of motion picture film in which display elements move realistically. This type of testing is then termed "dynamic open-loop" testing. A detailed description of this type of testing is given in Schum, Elam and Matheny (1962) in which its use has been demonstrated in connection with the study of design of altimeter is poor compared with other designs.

The point to be made is that human factors evaluation should begin at the point of assignment of functions in the development process and continue through to use in the fleet. During this evaluation process at least five methods of evaluation are applicable. There are (1) human engineering check lists, (2) static and dynamic open-loop tests, (3) mock-up evaluations, (4) tests in the simulator of the system and (5) tests in the actual system. A suggested relationship of these types of testing throughout the development and employment of a system is given in Figure 3.

Human engineering check lists are particularly appropriate to early stages of human engineering evaluation and can be used with the mock-up for type one evaluations. A listing of check lists felt to be representative of those in use as given in Section 8.0, Source Material. Open-loop testing lends itself to comparative evaluations of components and subsystems. The simulator and aircraft are most effective in evaluating whether or not human performance meets specific criteria for system effectiveness.

FIGURE 2.

POINT OF ENTRY OF OPEN-LOOP TESTING IN THE EVALUATION PROCESS

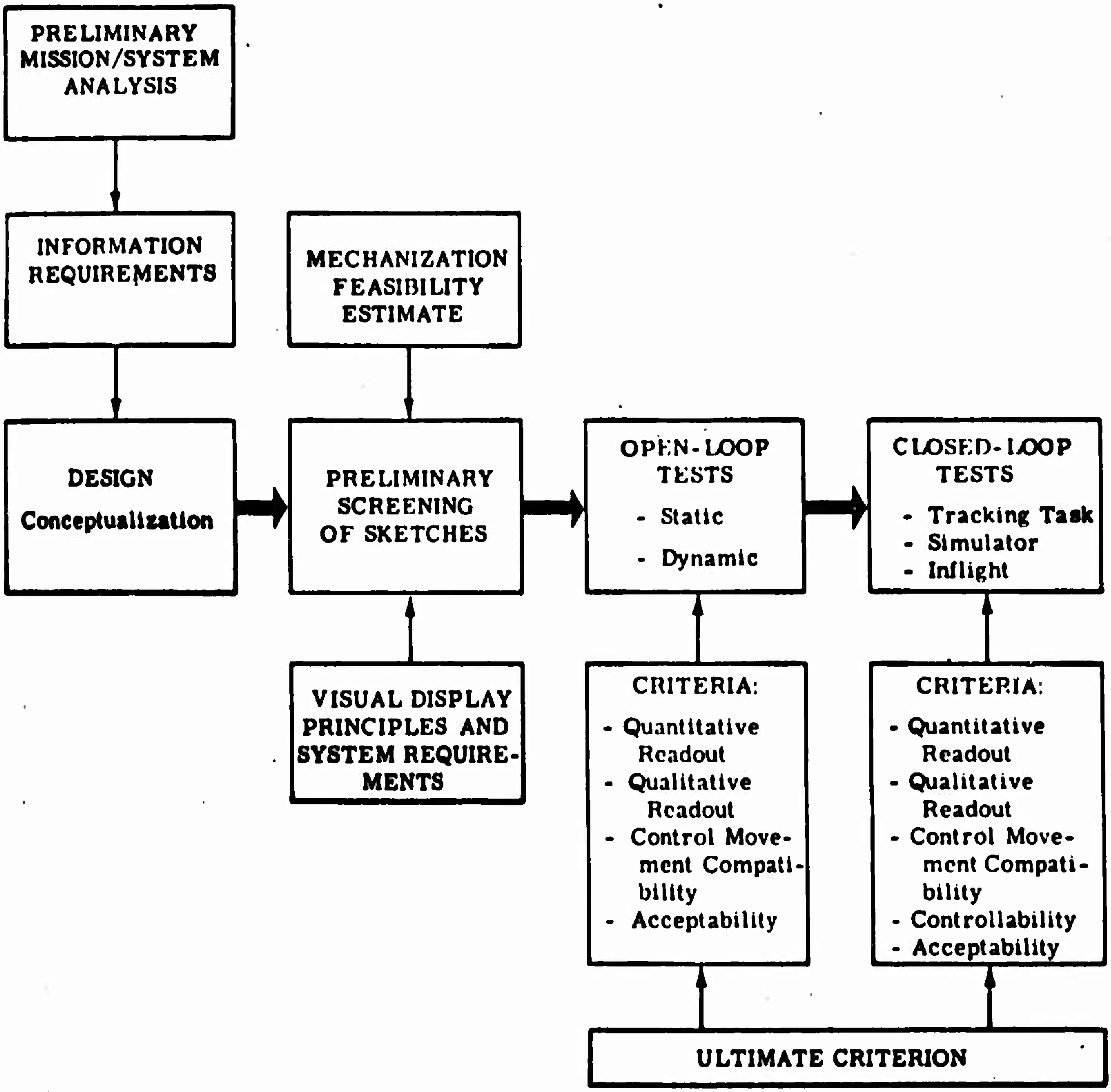
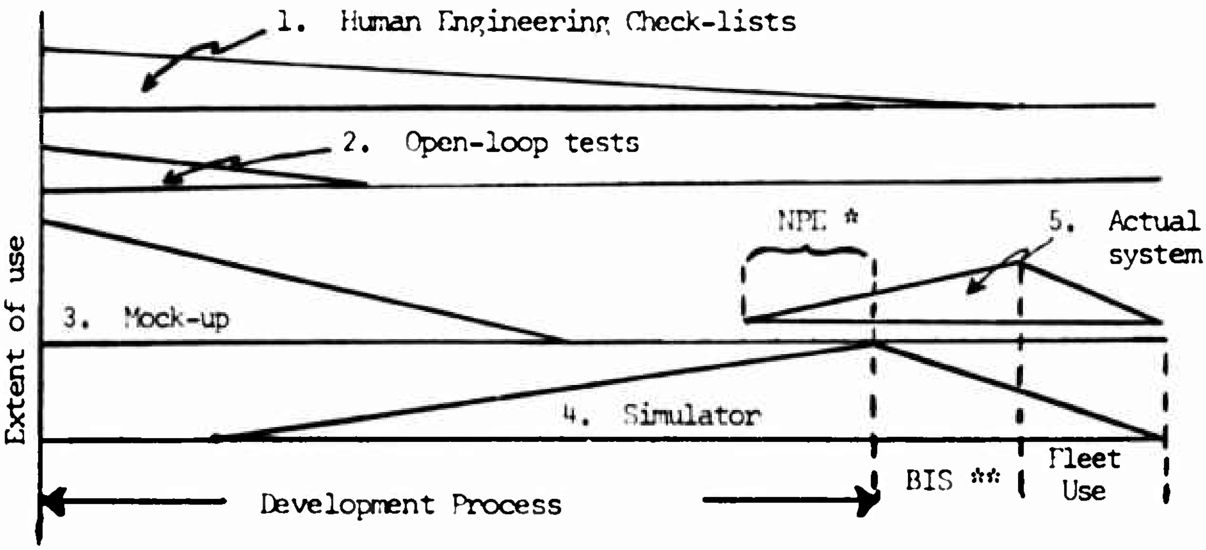


FIGURE 3.

RELATIONSHIP AND EXTENT OF USE OF EVALUATION PROCEDURES THROUGHOUT THE SYSTEM DEVELOPMENT PROCESS (SUGGESTED FOR STUDY)



1. Human engineering check-lists used to evaluate design compatibility with human capabilities and limitations.

2. Open-loop tests used for comparative evaluations of display and configuration design.

3. Mock-up used in conjunction with human engineering check-lists to determine design compatibility with human capabilities and limitations.

4. Simulator used in development phase to evaluate performance against system, subsystem and component performance criteria and to train Naval personnel. In BIS and Fleet Use it is used as in development plus:
   - o Evaluate proposed new or modified tactics
   - o Accident investigation
   - o Diagnosis of performance which is below criterion requirements

5. Actual system used to evaluate performance against system performance criteria and to evaluate proposed new or modified tactics.

* NPE - Navy Preliminary Evaluation

** BIS - Board of Inspection & Survey

Under this conceptualization the mock-up inspection as such takes on a different meaning. Under it the mock-up of the operator station as a design and evaluation tool is used on a continuing basis by manufacturer personnel under continuing monitorship of user personnel. The significance of and necessity for mock-up inspections are considerably reduced. As indicated earlier a greater stability in user monitoring personnel would be required and a closer and continual interaction effected between manufacturer and user evaluation personnel.

5.0 PERSONNEL AND ORGANIZATIONAL CONSIDERATIONS

5.1 DOCUMENTATION

In the recent past the requirements for consideration of human factors in the design and evaluation of Navy systems has become more explicitly documented. This comes about through adoption of MIL-H-46855, Human Engineering Requirements for Military Systems, Equipment and Facilities, 16 February 1968 and MIL-STD-1472, Human Engineering Design Criteria for Military Systems, Equipment and Facilities, 9 February 1968. These two documents when cited in a contract specification provide authority for carrying out effective human factors effort. Documentary authority without adequate supporting personnel to implement it at the managerial and working levels dilutes its effectiveness. The documents contain the authority for aggressive and definitive human factors work during RDT&E when enough human factors personnel are available. At the same time the documents themselves cannot spell out the requirements in the detail which ensures accomplishment of a good human factors effort in the absence of trained personnel in sufficient numbers dedicated to making their contribution felt.

5.2 MANAGEMENT

For human factors evaluation, assignment of the right personnel begins at the project office with responsibility vested in a designated individual for the evaluation function. It should be his responsibility to see that the system and time line analyses contain the information on test points and their priorities (see Section 3.2). These test points should be within the context of an evaluation mission. This mission set forth early in development will evolve in detail as the equipment configuration becomes firmed up. It is the framework upon which both Human Engineering design and evaluation will hang. The human factors engineer in the Project Office must insure that this mission and the evaluation test points are developed.

5.3 ASSIGNMENT OF NAVY PERSONNEL TO CONTRACTOR FACILITIES

The assignment of customer personnel to the human factors effort at the contractors facility has already been suggested. It is suggested that, optimally, these personnel be graduates of the Navy Test Pilot School with special additional training in human factors and recent experience in systems similar to the one being evaluated. The importance of the human component to the system and the scarcity of hard data on his performance in such systems warrants giving especial attention to the qualifications of the design and evaluation personnel who deal with him.

These Naval personnel should be selected for their interest in and any special qualifications for human factors work. Paralleling the practice of test pilots having the additional qualifications of holding aeronautical or other engineering degrees the human factors specialty should require special qualifications in human engineering. A special course of three to four weeks minimum should be made available for this specialty coming after the individual has qualified as a Naval Test Pilot. This special training course is discussed further in Section 5.5 and a suggested outline is given in Appendix B. This course is suggested as a minimum requisite and not a substitute for formal University training in Human Factors. The responsibilities of these personnel at the contractors facility would be purely advisory in helping to develop the evaluation mission and the critical test points. They would become members of the evaluation team during mock-up inspection lending their detailed knowledge of the operator requirements and the data flow through the system to a more objective and mission oriented evaluation.

5.4 CONTRACTOR ASSIST DURING NAVY EVALUATIONS

During evaluations carried out at the Navy's facilities it is recommended that contractor human factors personnel assist and advise on-site in the planning and conduct of the tests. These contractor personnel contribute their knowledge of the system to the planning and conduct of the tests. Their detailed knowledge will help the Navy evaluator immeasureably in working out the details of how to test at critical test points and in diagnosing sources of operator difficulties when his performance cannot be directly measured or clearly inferred.

5.5 TRAINING OF EVALUATION PERSONNEL

The approach to be taken by the evaluator in any evaluation is essentially that which the serious experimenter would take in testing an hypothesis. He must be as knowledgeable and have as much quantitative information as possible about the variables and conditions influencing the operator behavior. He must either control these or be able to assess their effects. He must also have an understanding of experimental design, of reliability of measurement and of data analysis and report.

An appreciation of these requirements coupled with experience and knowledge of the operational conditions under which the system will function would combine to maximinize the effectiveness of the human factors evaluator. Their combination in a single individual is rare. A training program designed to produce such a combination would be quite beneficial.

The human factors engineer usually comes to the evaluation situation with a limited knowledge and appreciation of the operational demands. The Navy project personnel assigned to evaluate the system have the operational experience but usually are not experienced in the methods of experimentation which should be applied. A cross training program is recommended.

It is suggested that Navy project personnel could become oriented and sufficiently knowledgeable about experimental methods through an indoctrination course of three to four weeks minimum. This course would be offered to project officers and project pilots who are directly concerned with planning and conducting the evaluation. A suggested outline for this course is contained in Appendix B.

Human factors personnel should be indoctrinated in the operational use of like systems in every way possible. Cdr. Wherry has suggested that short tours aboard carriers by human factors personnel should be undertaken. Every opportunity for these personnel to observe the operation of similar systems either in the operational theater or more routine operation should be taken.

6.0 CONCLUSIONS AND RECOMMENDATIONS

After reviewing relevant Navy documents and pertinent reports in the general literature, holding interviews and discussions with those experienced with and involved in design and evaluation, studying the design and evaluation process of one system in detail with a less detailed look at two other aircraft systems, and drawing upon first hand experience with the design and evaluation problem of a major aircraft system the conclusions and recommendations enumerated below have been drawn.

These recommendations stem from the philosophy that changes in the human factors evaluation procedure must be evolutionary rather than revolutionary. They must be evolutionary primarily because the required number of qualified people are not available to institute mandated revolutionary changes. Rather, people with the necessary orientation and qualifications must be integrated into the evaluation process and test organizations to evolve procedures and to develop facilities and personnel. The recommendations given here are intended to provide procedures, information and tools to support those individuals as they work directly with the evaluation of a system. They are not intended as "band-aids" designed as a temporary fix on the human factors evaluation process. Rather they point to fundamental needs. Some are innovative while some require doing more formally and systematically what is now done informally and even sporadically.

6.1 EVALUATION MISSION

A system mission should be described and adopted as the evaluation mission. This mission should be built around the primary weapons delivery mode as the system will operate in its most likely theater of operation and incorporate "worst case" conditions. (Reference Section 3.4)

6.2 TEST POINTS

The mission-centered system and time line analyses should produce test points at which human component performance may be measured directly or inferred reliably from equipment outputs. These test points must be ranked in terms of relative importance to mission success and extent to which assumptions were made as to operator capabilities. (Reference Sections 3.2 and 3.5)

6.3 HUMAN FACTORS AND EQUIPMENT ENGINEER COOPERATION

The measures to be taken at test points will often require the objective recording of certain measurement parameters. Often these parameters will be those being recorded by the design engineer in his evaluation of the system equipment. Through close coordination the design and the human factors engineers can use the same test equipment to obtain data peculiar to their own needs. The cooperative planning of human factors and equipment tests is recommended as essential to effective use of test equipment and test time. (Reference Section 3.7)

6.4 MEASURES OF ABSOLUTE LEVEL OF PERFORMANCE

Although the test measurements taken during a field evaluation are intended to determine whether performance meets required standards, it is strongly recommended that measures of absolute level of performance also be obtained whenever practicable. It is also recommended that complete descriptions of the parameters of the equipment from which the operator receives his inputs and into which he makes his outputs be given. The environmental and test conditions under which the measures were taken and the level of training and experience of the operator must also be completely described. These data are necessary for building a pool of data for use during design of future systems (Reference Sections 3.3 and 4.3.1)

6.5 TROUBLE SHOOTING

Formal reporting procedures and forms are recommended for operator-evaluator use in identifying operator errors or failure to perform to the required time schedule during the test of subsystems in which the operator performance cannot be measured directly or inferred reliably. (Reference Section 4.2.2)

6.6 ADOPTION OF ALTERNATIVES BY THE OPERATOR

Formal reporting procedures and forms are recommended for operator-evaluator use in reporting instances in which it was necessary for him to adopt a procedure or technique different from the design procedure or technique in order to successfully accomplish his task or mission.

6.7 MISSION ORIENTED COCKPIT EVALUATION

In working with the cockpit mock-up evaluation, procedures and check-lists are recommended which are system-mission oriented in addition to those which are cockpit evaluation oriented. (Reference Section 4.3.2)

6.8 DYNAMIC PHYSICAL MODELS (SIMULATORS)

A dynamic physical model of the system (simulator), if made available during early design would allow for evaluation with a much more representative model of the final aircraft than is possible with the static mock-up. It is recommended that the advantages and disadvantages of a major simulator installation at a Naval Test Facility be investigated in detail. These investigations should set forth the pros and cons of a facility designed for a single and dual seat aircraft, degree of flexibility required and its applicability to design evaluation and refinement throughout the development process. (Reference Section 4.3.2)

6.9 ASSIGNMENT OF NAVAL PERSONNEL TO CONTRACTORS FACILITY

It is recommended that the mock-up inspection procedure be strengthened by the assignment to the contractors facility of customer personnel especially trained in human factors evaluation and experienced in prior systems. These personnel would work with human factors personnel in defining the evaluation mission, test points and evaluative procedures. These personnel, although working in an advisory capacity during design, would become members of the mock-up inspection team during that evaluation (Reference Section 5.3)

6.10 ASSIGNMENT OF CONTRACTOR PERSONNEL TO FIELD EVALUATIONS

It is recommended that contractor human factors personnel work with Naval evaluation personnel during field evaluation of the system. Contractor personnel would provide system information and assist in evaluation planning. (Reference Section 5.4)

6.11 SPECIALIZED TRAINING PROGRAMS

It is strongly recommended that special training in human factors design and evaluation methods be provided to Navy personnel responsible for human factors test and evaluation. A three to four week course after Navy Test Pilot qualification is recommended as a minimum requisite but not intended to substitute for formal University training in Human Factors.

It is also recommended that evaluation personnel whose speciality and background of training is in Human Factors but who do not have pilot and/or operational experience be given indoctrination tours with the fleet to observe the operation of similar systems and to discuss operational tactics and problems with fleet personnel. (Reference Section 5.5)

7.0 APPLICATION OF CONCLUSIONS

This study of the human factors field evaluation problem has resulted in conclusions and recommendations deduced from the general literature, Navy documents, examination of case histories, interviews and the investigator's own experience with the problem. These conclusions and recommendations, to be useful to the Navy, need demonstration of their workability within the field test and evaluation context. To this end the next step in this investigation will be the trial of certain of the recommendations during the field test of a specific system.

The system within which recommendations will be tested is the P-3C aircraft. The system analyses will be examined to determine test points and their priorities. Performance measures and methods of data recording will be specified and data recording will be undertaken in the field. From these activities the details of the field test procedure will be evolved and conclusions and recommendations will be modified as necessary.

8.0 SOURCE MATERIAL

8.1 SYSTEMS ANALYSIS AND INFORMATION REQUIREMENTS

Bloom, Joel N., et al. People, organizations and communications, Phase I. A model for communications system analysis, Army Communications System Division, Office of the Chief Signal Officer, Headquarters Department of the Army, June 1960.

Havron, M. Dean and Jenkins, J. P. Information requirements methods as applied to development of displays and consoles., Engineering Psychology Branch, Psychological Sciences Division, Office of Naval Research, HSR-RR-61/4-Sm, March 1961.

Hitchcock, L. Jr. The analysis of human performance within the operational flight environment, in Assessment of Skill and Performance in Flying, AGARD Conference Proceedings, No. 14, September 1966.

Kahn, H., et al. Techniques of systems analysis, The Rand Corporation, Santa Monica, California, December 1956.

Kurke, M. I. Operational concept analysis and sources of field data, Human Factors, Vol. 7, No. 6, December 1965.

McGrath, J. E. and Nordlie, P. G. Synthesis and comparison of system research methods, Human Sciences Research, Inc., Arlington, Va., Report Number 9, Contract Nonr 2425-(00), February 1960.

Meister, David and Rabideau, G. F. Human Factors Evaluation in System Development, John Wiley, New York, 1965.

Porter, E. H. A paradigm for system analysis of command and control functions, System Development Corporation, Santa Monica, California, October 1964.

Wilson, Clark L. On-the-job and operational criteria, in Glaser, R., Training Research and Education, University of Pittsburgh, prepared for Office of Naval Research under Contract Nonr 624(09). Page 410-412.

8.2 HUMAN ENGINEERING CHECKLISTS

Cornell Aeronautical Laboratory, Inc. Pocket data for human factors engineering, June 1959.

Fitzpatrick, R. A checklist for human factors in operational suitability tests, American Institute for Research, Contract AF 24(601)-124, Project No. 7800, Research Directorate, Air Force Special Weapons Center, Air Research and Development Command, Kirtland AFB, New Mexico, August 1955.

Greek, D. C. Checklist of human engineering design principles, MD58-334, North American Aviation, Inc., Missile Division, January 1959.

Meister, David and Rabideau, G. F. Human Factors Evaluation in System Development, John Wiley, New York, 1965.

Philco-Ford Corporation, Human Engineering Design Check List, WDL-TR1968A, Philco-Ford Corporation, WDL Division, Palo Alto, California, 8 May 1964.

Plath, D. W. Human engineering checklist. Autonetics, A Division of North American Aviation, Inc., January 1966.

8.3 PERFORMANCE EVALUATIONS AND SYSTEM EFFECTIVENESS

Adams, Jack A. and McAbee, W. H. A program for evaluation of human factors in category II testing of air weapons control system 412L (phase II configuration). Air Force Systems Command, United States Air Force, Eglin Air Force Base, Florida, May 1962.

Air Force Ballistic Missile Division (ARDC), Personnel subsystem testing for ballistic missile and space systems, AFBMD Exhibit 60-1, April 1960.

Barfoot, C. B. A Preliminary Cost-Effectiveness Handbook. Headquarters, United States Army Combat Developments Command, Ft. Beivoir, Va., 15 December 1963.

Barrett, G. V., Thornton, C. L. and Cabe, P. A. Human factors evaluation of a computer based information storage and retrival system, Human Factors, Vol. 10, No. 4, August 1968.

Borchers, Kenneth H. Are ballistic missile test programs structured to support adequate evaluation of human performance? Human Factors, Vol. 6, No. 6, December 1964.

Borden, Gail J. and McGrath, James J. Geographic orientation in aircraft pilots: Field validation of a post-flight method of reporting navigation performance, JANAIR Report 680714, Human Factors Research, Incorporated, Santa Barbara Research Park, Goleta, California, July 1968.

Bowen, H. M., et. al. Study, assessment of pilot proficiency. U.S. Naval Training Device Center, Port Washington, N. Y., Technical Report: NAVTRADEVCEN 1614-1, August 1966.

Bowen, H. M., et. al. Studies of divers performance during the sealab II project, Office of Naval Research, Washington, D. C., Contract No. Nonr 4930(00), March 1966.

Brictson, C. A. Measures of pilot performance: Comparative analysis of day and night carrier recoveries. Dunlap and Associates, Inc. Santa Monica, California, June 1966.

Bureau of Naval Personnel, Human factors prediction evaluation, Prepared by Data-Design Laboratories for the Bureau of Naval Personnel, Contract Nonr 4739(00), August 1967.

Chapin, P. W., Detrick, R. E. and Ilg, R. P. Flying qualities and performance evaluation of the A-7A airplane, Report No. NATC-FT-35R-68, May 1968, Naval Air Test Center, Patuxent River, Maryland, Project A055 100 40, Task A530111-6.

Cooper, L. J. and Manion, R. C. Testing the OPTEVFOR Guidebook for the collection of human factors data, Report 1: Direct Observation, Report 4040F-1, HRB-Singer, Inc., State College, Penn. November 15, 1967, (Preliminary Report).

Curtin, J. G., et al. Evaluation of the RH-2 cockpit mock-up. Bell Helicopter Company, Fort Worth, Texas, August 1965.

Glaser, R. Evaluation of personnel proficiency, in Human Factors Methods for Systems Design. John D. Folley, Jr., Editor, AIR-290-60-FR-225, American Institute for Research, 1960, under Contract No. Nonr 2700(00), AD 232 646.

Gross, R. L., Lindquist, O. H., Peterson, J. R. & Blanchard, J. W. The application, validation and automation of a method for delineating and quantifying aerospace flight crew performance, in Symposium and Workshop on the Quantification of Human Performance at Albuquerque, the University of New Mexico, 1964.

Hagen, W. C. Techniques for the allocation and evaluation of human resources, OR 8735, Martin Marietta Corporation, Orlando, Florida, March 1967.

Houston, R. C., Smith, J. F. and Flexman, R. E. Performance of student pilots flying the T-6 aircraft in primary pilot training, 1954, AFPTRC-TR-54-109.

Jeantheau, G. G., Andersen, B. G. Training system use and effectiveness evaluation, U. S. Naval Training Device Center, Port Washington, N. Y., Technical Report: NAVTRADEVCEN 1743-1, June 1966.

Johnson, Dr. G. W. Navy laboratory role in human performance effectiveness, in Proceedings, Symposium on Human Performance Quantifications in Systems Effectiveness, sponsored by Naval Material Command and the National Academy of Engineering, Washington, D. C., January 1967 AD 650 791.

Keenan, J.J., Treadway, C. P. and Lenzycki, H. P. Concepts and practices in the assessment of human performance in Air Force systems, AMRL-TR-65-168, Aerospace Medical Research Laboratories, AFSC, Wright-Patterson AFB, Ohio, September 1965.

Kryter, K. D. Proposed methods for the calculation of the articulation index, ESD TDR-62-5, Operational Applications Office, Electronics Systems Division, L. G. Hanscom Field, Bedford, Mass., October 1961.

Kryter, K. D., Flanagan, G. and Williams, C. A test of the 20-band and octave-band methods of computing the articulation index, ESD-TRD-62-4, Operational Applications Laboratory, Electronic Systems Division, L. G. Hanscom Field, Bedford, Massachusetts, October 1961.

Kryter, Karl D. and Ball, J. H. SCIM -- A meter for measuring the performance of speech communications systems, TDR Report No. ESD-TDR-64-674, Decision Sciences Laboratory, Electronic Systems Division, Air Force Systems Command, October 1964.

Landix, D., et al. Experiments in display evaluation, Phase I report, Technical Report 1-194, The Franklin Institute Research Laboratories, Contract Nonr-4832, July 1967.

Leuba, Boteilho, R. J. Evaluation of system design adequacy., ARINC Research Corporation, a subsidiary of Aeronautical Radio, Inc., Washington, D. C., December 1963.

Luehrs, R. E., Capt., MC, USN, Human error research and analysis program (HERAP), in Assessment of Skill and Performance in Flying, AGARD Conference Proceedings No. 14, September 1966.

Lytton, L. E. Evaluation of a vertical-scale, fixed-index instrument display panel for the X-15 airplane, NASA TN D-3967, National Aeronautics and Space Administration, Washington, D. C., May 1967.

McKendry, J. M. and Harrison, P. W. Assessing human factors requirements in the test and evaluation stage of systems development, Volume 1., HRB-Singer, Inc., State College, Pennsylvania, Contract Nonr 4203(00), June 1964.

McKendry, J. M., et al. Assessing human factors requirements in the test and evaluation stage of systems development, Vol. 11, HRB-Singer, Inc., State College, Pennsylvania, Contract Nonr 4203(00), June 1964.

Mahler, W. R., et al. Psychological studies of advanced Naval air training: Evaluation of operational flight trainers, Psychological Corporation, New York, September 1950.

Manion, R. C., Cooper, L. J. and Thomas D. B. Testing OPTEVFOR guidebook for the collection of human factors data, Report 2: The questionnaire, Report 4040F-3, HRB-Singer, Inc., State College, Penn. November 1967 (Preliminary Report).

Manion, R. C. and Cooper, L. J. Testing the OPTEVFOR guidebook for the collection of human factors data, Report 3: The checklist, Report 4040-F-2, HRB-Singer, Inc., State College, Penn. November 1967 (Preliminary Report).

Manion, R. C. and Cooper, L. J. Testing the OPTEVFOR guidebook for the collection of human factors data, Report 4: The interview, Report 4040F-4, HRB-Singer, Inc., State College, Penn. November 1967 (Preliminary Report).

Markel, G. A. Toward a general methodology for systems evaluation, HRB-Singer, Inc., State College, Pennsylvania, July 1965.

Meister, David. The measurement of man-machine systems under field operational conditions, Paper presented at the ONR Ninth Annual Human Engineering Conference, St. Louis, Missouri, June 1-2, 1961.

Meister, David. The measurement of man-machine systems under field operational conditions, Human Factors, Vol. 3, No. 4., December 1964.

Meister, David. Methods of predicting human reliability in man-machine systems, Human Factors, Vol. 6, No. 6, December 1964.

Meister, David and Rabideau, G. F. Human Factors Evaluation in System Development, John Wiley, New York, 1965.

Newell, F. D. and Pietrzak, P. E. In-flight measurement of human response characteristics, AIAA Flight Test, Simulation and Support Conference, Cocoa Beach, Florida, February 6-8, 1967.

Obermayer, R. W. Simulation, models and games: Sources of measurement, Human Factors, Vol. 6, No. 6, December 1964.

Pickrel, E. W. and McDonald, T. A. Quantification of human performance in large, complex systems, Human Factors, Vol. 6, No. 6, December 1964.

Rabideau, G. F. Field measurement of human performance in man-machine systems, Human Factors, Vol. 6, No. 6, December 1964.

Rappaport, Maurce. Further considerations about the what and how of field testing, Unpublished report. 1961.

Rook, Luther W., Jr. A method for evaluating the human error contribution to system degradation, Human Error Quantification. A Symposium, A. D. Swain, Chairman, SCR-610, Sandia Corporation, Albuquerque, New Mexico, April 1963.

Rook, Luther W. Jr. Evaluation of system performance from rank-order data, Human Factors, Vol. 6, No. 5, October 1964.

Rook, Luther W. Jr. Evaluation of system performance from rank-order data, Human Factors, Vol. 6, No. 5., October 1964.

Schum, David A., Elam, Claude B. and Matheny, W. G. Altimeter display study - Handbook for static testing, Bell Helicopter Company, Report, Contract AF33(616)-8236, Project 6190, Task 82144, Aeronautical Systems Division, U. S. Air Force Systems Command, Wright-Patterson AFB, Ohio, January 1962.

Shearer, J. W., Peterson, D. A., and Slebodnick, E. B. Techniques for human factors evaluation of prototype special weapons and associated equipment, American Institute for Research, Contract AF 29(601)-513, Air Research and Development Command, Kirtland AFB, New Mexico, 1959.

Sheldon, M. S. and Zagorski, H. J. Project NORM Field Evaluation Test Report, System Development Corporation, Santa Monica, California, Contract AF 19(628)-5166, July 1966.

Siegel, Arthur I. and Wolf, J. Jay. Techniques for evaluating operator loading in man-machine systems: A description of a model and the results of its first application, Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, February 1959.

Siegel, Arthur I. and Wolf, J. Jay. Techniques for evaluating operator loading in man-machine systems: Application of a previously derived model to the launching of an air-to-air missile, Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, June 1959.

Siegel, Arthur I., et al. Techniques for evaluating operator loading in man-machine systems: A "model" for digital simulation of one and two-operator man-machine system, Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, March 1961.

Siegel, Arthur I. and Wolf, J. Jay. A technique for evaluating man-machine system designs, Human Factors, Vol. 3, No. 1, March 1961.

Siegel, Arthur I. and Wolf J. Jay. Techniques for evaluating operator loading in man-machine systems: A further application of a "model" for digital simulation of one or two-operator man-machine systems, Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, June 1961.

Siegel, Arthur I. et al. Techniques for evaluating operator loading in man-machine systems: Evaluation of a one or a two-operator system evaluative model through a controlled laboratory test, Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, July 1962.

Siegel, Arthur I. et al. Absolute scaling of job performance, Journal of Applied Psychology, 1968, 52, (4) 313-318.

Swain, Alan D. Some problems in the measurement of human performance in man-machine systems, Human Factors, Vol. 6, No. 6, December 1964.

Smode, A. F., Gruber, A., and Ely, J. H. The measurement of advanced flight vehicle crew proficiency in synthetic ground environments, Report MRL-TDR-62-2, Behavioral Science Laboratory, Aerospace Medical Division, Wright-Patterson AFB, February 1962.

Ware, Claude T., Jr. Individual and situational variables affecting human performance, Human Factors, Vol. 6, No. 6, December 1964.

Widrewitz, J. Concepts relative to system effectiveness. Reliability Branch, Rome Air Development Center, Research and Technology Div., Air Force Systems Command, Griffiss Air Force Base, N. Y., June 1964.

William, Carl E., Hecker, Michael, H. L., and Kryter, Karl D. Methods for psychoacoustic evaluation of speech communication systems, Technical Documentary Report No. ESD-TDR-65-153, Decision Sciences Lab., Electronic Systems Division, Air Force Systems Command, December 1964.

Wilson, Clark L. On-the-job and operational criteria, in Glaser, R., Training Research and Education, University of Pittsburgh, prepared for Office of Naval Research under Contract Nonr 624(09). Page 410-412.

8.4 MODELS AND METHODS

Bloom, Joel N., et al. People, organizations and communications, Phase I. A model for communications system analysis, Army Communications System Division, Office of the Chief Signal Officer, Headquarters Department of the Army, June 1960.

Cagley, E. M. A general network simulation model, the U. S. Army Engineer Mathematical Computation Agency, Corps of Engineers, NREC Technical Report No. 31, April 1964.

Guilford, J. P. Psychometric Methods, (Second edition), New York, McGraw-Hill, 1954.

Hendricks, J. W. A system for development of mathematical models, U. S. Naval Radiological Defense Laboratory, San Francisco, California, USNRDL-TR-749, April 1964.

Leuba, H. R. Quantification in man-machine systems, Human Factors, Vol. 6, No. 6, December 1964.

Myers, Louis B., et al. Guidebook for the collection of human factors data (Second Printing), HRB-Singer, Inc., State College, Pennsylvania, Contract Nonr 4751 (00), June 1966.

Rook, Luther W. Jr. Evaluation of system performance from rank-order data, Human Factors, Vol. 6, No. 5, October 1964.

Rosen, S. B. et al. A study of system modeling techniques (Final Report), Air Force Systems Command, United States Air Force, L. G. Hanscom Field, Bedord, Mass., Contract NrAF 19(628)-2973, October 1963.

Siegel, Arthur I. and Wolf, J. Jay. Techniques for evaluating operator loading in man-machine systems: A description of a model and the results of its first application, Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, February 1959.

Siegel, Arthur I. and Wolf, J. Jay. Techniques for evaluating operator loading in man-machine systems: Application of a previously derived model to the launching of an air-to-air missile, Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, June 1959.

Siegel, Arthur I., et al. Techniques for evaluating operator loading in man-machine systems: A "model" for digital simulation of one and two-operator man-machine system, Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, March 1961.

Siegel, Arthur I. and Wolf J. Jay. Techniques for evaluating operator loading in man-machine systems: A further application of a "model" for digital simulation of one or two-operator man-machine systems, Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, June 1961.

Siegel, Arthur I. et al. Techniques for evaluating operator loading in man-machine systems: Evaluation of a one or a two-operator system evaluative model through a controlled laboratory test, Applied Psychological Services, for Engineering Psychology Branch, Office of Naval Research, July 1962.

Terry, R. A. et al. Adaptive processes used in effects modeling. Presented at the first symposium on The Psychological Effects of Non-Nuclear Weapons, Eglin Air Force Base, Florida, April 1964.

Torgerson, Warren S. Theory and Methods of Scaling, (Second Printing) New York, John Wiley & Sons, 1960.

8.5 DATA STORES (BANKS)

Barrett, G. V., Thornton, C. L. and Cabe, P. A. Human factors evaluation of a computer based information storage and retrival system, Human Factors, Vol. 10, No. 4, August 1968.

APPENDIX A

Reference Table of Measurement Parameters Applicable to Operator Performance Evaluation in Aerospace Craft.
(For a detailed discussion of measurement parameters see text, Section 3.7, Pages 14 - 19).

ABOUT AXIS CONTROL
(Attitude)

Pitch angle
Roll angle
Yaw or heading angle
Pitch rate (q)
Roll rate (p)
Yaw rate (r)
Control position
Control rate

ALONG AXIS CONTROL
(Navigation)

Z Axis position (alt.)
X axis position
Y axis position
Vertical rate ($\dot{Z}$)
Longitudinal rate ($\dot{X}$)
Lateral rate ($\dot{Y}$)
Vertical acceleration ($\ddot{Z}$)
Lateral acceleration ($\ddot{Y}$)

"SYSTEM STATE" CONTROL

Proper control position and display reading and execution of proper procedures for subsystems:

- Weapons
- Communications
- Hydraulic
- Environmental
- Electrical
- Flight Controls
- Engine and related systems
- Fuel
- Radar

APPENDIX B

This appendix contains the suggested outline of a course designed for field test personnel whose particular interest and concern is human factors testing. The course is intended to cover items essential to an overall orientation in human factors test and evaluation and to provide specific guidance in methods and techniques necessary to effective evaluation.

In this course outline a differentiation is made between field research and field evaluation with similarities and differences drawn between them. While the primary aim of the course is to provide guidance for field evaluation, points relevant to field research are given in order that the human factors evaluation may be sensitive to and have some guidance in field research techniques should the opportunity arise to apply them during a field evaluation. His motive for doing so would be the collection of data under prescribed conditions for entry into the general human factors data bank.

COURSE OUTLINE

1. Distinction between field research and field evaluation.

   A. Field Research

      1) More opportunity to identify and control the parameters and variables of interest.

      2) Usually collecting normative data.

      3) May be testing to determine how well system, subsystem or component meets some set or required level of performance.

      4) More opportunity to introduce performance measures and special instrumentation to obtain those measures.

      5) More flexibility in changing procedures and equipment as testing progresses to achieve the desired goals of testing.

   B. Field Evaluation

      1) Nearly always testing to determine whether system performance comes up to some specified level of desired performance.

      2) Nearly always there is a limitation on time within which the evaluation is to be performed.

      3) Must fit measures of man's performance into the tests of equipment.

4) Need diagnostic measures to isolate those points at which performance has failed to meet criteria when total system performs below criterion level.

5) Little or no opportunity to vary independent variables systematically.

2. General setting within which field research and evaluation are conducted.

A. Some fixed time span.

B. Testing both equipment and men.

1) Some tests peculiar to equipment alone, e.g., how it functions under the field conditions.

2) Some tests peculiar to man alone, e.g., his physiological state under the field condition.

3) Some tests peculiar to the interaction between men and equipment, e.g., how well man can operate or maintain equipment under the field conditions.

C. At present human factors testing often goes "piggy back" on equipment testing -- or must be fitted in and around equipment testing. This is a fact of human factors testing which must be recognized in setting up human factors research and evaluations in the field. This situation is not to be construed as all bad. A great deal of human factors data collecting can be carried out in conjunction and simultaneous with equipment testing. It is the efficient and many times the most appropriate thing to do. However, it will be necessary at times to program into the testing schedule specific blocks of time for collecting human factors data independent of tests on any given piece of system equipment.

3. The requirement for thorough knowledge of the system under test, its operation and its operating environment.

- To isolate and define the important parameters which may influence performance.
- To set up methods of either controlling or systematically varying these parameters.
- To determine what measures are appropriate and at what test points they will be taken.

4. Methods for determining necessary system details.

A. Data flow analysis.

1) What inputs are (or must be) received by each component or subsystem and what outputs are (or must be) made to the next component(s) or subsystem(s).

2) For man as an information processor we are interested in the following general types of information: a) what information must be receive or is the system designed for him to receive? b) what transformations of the information must he make? c) what "informational" outputs must he make to other components of the system either man or machine?

3) Nomenclature or symbology used is not necessarily important however those which may be used are.

B. Time Line Analysis (serves both as a means of learning the system and evaluating it - also serves to evaluate the adequacy and practicality of the test and evaluation proposed to be carried out.)

1) Segment overall program or mission into phases which can be bounded in time by externally imposed limits.

2) Detail actions and procedures to be executed within each segment.

3) Assign "time to perform" to each action or procedures obtaining times either from available data, estimates or from data (taken in mock-up simulator or actual system.)

4) Compare actual times to complete segments to time required as given by program or mission requirements given in 1.

C. A list of techniques for system operational description and a exposition of the Operational Sequence Diagram is given in "Human Factors Design Standards for the Fleet Ballistic Missile Weapon System: NAVWEPS OD 18413A, Vol. 1, Pages 50-60.

5. Some general principles of sound research and evaluation procedure for which to aim.

A. Standardization

1) Test conditions.

2) Performance measures.

3) Observers.

4) Environmental effects.

B. Control or assessment of relevant variables.

1) If variable cannot be controlled it should be measured at the time dependent variables are measured.

C. Explicit statement of independent and dependent variables with no variation in their meaning or method of measurement.

D. Dependent variables.

1) These are, in effect, precise statements of what we want to know about how the system or the man within the system performs.

2) In field research we may term these questions the dependent variables. In field evaluation we may term them the criteria against which we measure system or man's performance.

3) Must order dependent variables in order of priority of importance since and/or time is usually limited.

4) Must decide upon points in system at which measurements will be taken.

5) Must determine the method of measurement, i.e., direct recording, direct observation, rating scales, questionnaire, etc.

6) Always detail in advance the method of data reduction, analysis and presentation. Shotgun approach to data collection is not feasible in field situations.

7) General classes of dependent variables.

a) Man's outputs

a. Time to perform.

b. Accuracy of performance.

b) Man's inputs.

a. Standard deviation of control displacements (or pressures) about the mean control position.

b. Power spectrum or autocorrelation function.

c) Man's physiological state.

d) Man's psychological state.

E. Independent Variables.

1) In field research may have the opportunity to assess, control or vary systematically the independent variables. In field evaluation usually have opportunity only to assess these variables.

2) From system and task analysis identify and describe both system and environmental parameters like to affect performance.

3) Field research will normally be set up for the test of the effects of specified parameters upon performance.

4) General normative data vs. data specific to a system influence selection of independent variables.

6. Criteria

- Implies some value judgment as to the "goodness" of the performance.
- Measurement per se does not provide value judgments.
- These value judgments must be expressed in terms of the desired purposes or the mission of the system.

A. Ultimate vs. Actual Criteria

1) Seldom possible to obtain direct measures of the ultimate criteria.

2) Usually necessary to select some actual (intermediate) criteria.

3) Must then use these actual (intermediate) criteria in evaluating performance.

4) There is no certain method for specifying the actual criteria.

5) Sources of error in selecting actual criteria.

a. Unreliability.

b. Irrelevancy - the lack of relation to ultimate criterion.

c. Contamination - ingredients in the actual criteria which do not, in fact, comprise the ultimate criterion.

d. Distortion - errors arising from assigning incorrect weights to the separate factors that comprise the actual criteria.

B. Establishing Valid Criteria.

1) No established procedures.

2) Recognizing importance of criteria selection and types of errors which might be present are good starting points.

3) Steps which should lead to more useful and relevant criteria.

   a. Define the activity - specify to extent possible the activity desired for successful and proficient performance.

   b. Analyze the activity - consider the activity in terms of purposes or goals, behavior and skills involved, their relative importance and standards of performance expected.

   c. Define proficient and successful performance.

   d. Develop sub-criteria to measure each element of success.

   e. As appropriate develop a combined measure of successful performance.

C. Combining Actual Criteria.

1) Often necessary that several criteria, all of which are relevant for a particular activity be used. In such case it may be desirable to combine them into a single comprehensive one.

2) Combining will usually involve assigning relative weights to the individual criteria.

3) Rules for combining criteria.

   a. Weight in accordance with their relevance to the ultimate criterion.

   b. Criteria which repeat or overlap factors in other criteria should receive low weight.

   c. Other things being equal the more reliable criteria should receive more weight.

4) Caution must be exercised in applying weights to raw score values - use standard scores.

7. Measurement of Performance.

A. What to measure.

1) Should be preceded by an explicit statement of the research questions being asked.

2) Even though one is carrying out field research only to obtain normative data, the experimental questions are couched within the framework of some background knowledge of systems and tasks about which data is needed to determine:

a) Whether man can perform certain tasks and to what level of performance.

b) How well the system would function if man were placed as a component within the system.

3) Need analyses of systems which allow statement of the critical tasks within those systems or classes of systems about which data is needed.

4) Identification of critical tasks.

a) What is measured depends on purpose so procedures and emphasis will be different for research, for training and for proficiency evaluation.

b) Not necessary or feasible to measure everything - rather measure a sample of behavior. Shotgun approach to measurement not desirable.

c) As a rule, select for measurement those tasks on which good performance leads to mission success and poor performance leads to mission failure.

d) In identifying and defining tasks and mission segments critical to system performance start with a descriptive analysis on a time line basis of the tasks that make up system operation.

e) In identifying critical tasks asking the following questions with respect to the tasks is helpful.

Would below-minimum performance;

- lead to an accident?
- result directly in mission failure?
- be impossible to remedy within the time constraints or not at all?
- be difficult to detect because of inadequate information feedback?
- recur over time in such a way as to produce a cumulative effect?
- contribute a large proportion of time to the total time required for some larger and critical function?

f) Emphasis to be placed on individual activities and crew-conducted activities.

1) Should be determined from an analysis of the relative importance of the two within the system.

2) Will depend on the stage of learning - ordinarily, individual activities are stressed in earlier learning stages whereas crew functions are emphasized in terminal phases of training.

B. Levels of Measurement

1) Nominal Scale.

a. Provides identity only.

b. Assigns unit or event to class or set.

2) Ordinal Scale

a. Provides both identity and order.

b. The units are assigned a rank order.

c. Does not connote quantitative measurement as such but rather a judgment of the amount possessed by the units involved.

3) Interval Scale.

a. Provides identity, order and additivity.

b. Units are scaled in equi-distant terms.

c. Intervals between quantities are equal.

d. Zero point is arbitrary and does not designate complete absence of the property.

4) Ratio scale.

a. Provides identity, order and addivity with reference to an absolute zero.

b. An extension of the interval scale with a natural absolute zero point.

C. Purpose of Measurement.

1) Prediction of future success.

2) Evaluation of present performance.

3) Evaluation of learning rate.

4) Identification of areas of strength and deficiency.

5) Evaluation of training effectiveness.

6) Selection and placement of individuals and teams.

7) Refinement of criterion information.

8) Definition of requirements.

D. Overall vs. Diagnostic Measures

1) Over-all Measures.

a) Global indices of sub-system or system performance associated with mission segments or complete mission.

b) Useful in assessment since it is descriptive of some end result which can be compared with the standard.

c) Weak in analytic sense since they provide no detailed information on performance beyond the outputs sampled.

2) Diagnostic Measures.

a) Quite specific, identifying elements of job performance in specific skill areas.

b) Since they are concerned with smaller more precisely defined units of behavior they lend themselves more readily to objective valid measurement.

c) Those relevant to criterion performance (predictive) are more useful in differentiating among operators or crews since terminal measures often contain uncontrolled variables.

E. Accuracy of Measurement.

1) Refers to how close the obtained value or measure is to be true value.

2) There is no single way to assure measurement accuracy - accuracy may be improved by the following means.

a) Increase scope of measurements to be taken - include additional aspects of relevant behavior.

b) Increase the number of observations on which means, etc. are based.

c) Control the conditions under which measurements are taken.

1) Define those factors present and held constant and those factors to be varied.

2) Maintain conformity of conditions throughout period of measurement.

3) Insure that the measures are taken correctly.

F. Reliability of Measurement.

1) Definition - agreement or consistency of measures from repeated observations.

2) Must have high reliability to obtain validity - may have high reliability but no validity.

3) Absolute expression of reliability.

a) Standard error of measurement - specifies the limits within which an obtained value may be expected to vary.

$$SE = s\sqrt{1 - r}$$

where: s = the standard deviation of the measures

r = the reliability of the measures expressed in terms of a correlation coefficient.

It is interpreted in a manner similar to the SD.

4) Relative measures of reliability.

a) Expressed in terms of correlation coefficient.

b) Coefficient of internal consistency, i.e., split half method.

c) Coefficient of stability - extent to which measurements agree over a period of time. Correlation between measures taken at identical observation points with an intervening time lapse.

d) Coefficient of equivalence - correlation between two different measures which are known or presumed to be equivalent.

e) Percent agreement of values taken during repeated observations.

G. Validity of Measurement.

1) Definition - degree to which measuring instruments measure what they are intended to measure.

2) Four types of validity.

a) Content validity - logical validity based on expert opinion or other logical considerations.

b) Concurrent validity - statistical validity - correlation with other task or dimension external to the measurement.

c) Predictive validity-statistical correlation between obtained measures and future states on some task or dimension external to the measurement.

d) Construct validity - logical validity - where the emphasis is on the trait, quality or ability presumed to underlie the measures being taken.

H. The question of Quantitative (Objective) vs. Qualitative (subjective) Measures.

1) Objective Measures.

a) Generally permit measurement relatively independent of the observer.

b) Generally of higher reliability than subjective.

c) Greatest objectivity obtained by means of recording instruments etc. where a permanent record of behavior is obtained at the time of occurrence.

d) Insistence upon complete objectivity tends to result in omission of a variety of critical job components because of inability to measure them objectively.

e) Can result in impractical gadgetry and procedures.

f) Relatively free from observer bias.

2) Subjective Measures.

a) Generally dependent upon the characteristics of the observers - may introduce bias.

b) Inter-observer reliability not always high.

c) More flexibility in administration.

3) Ratings (a form of subjective measurement)

a) Because of reliability considerations ratings should be reserved for use in those instances where other measures are not feasible.

b) Rating procedures.

1) Rating scales - rater makes judgment on scale of defined categories.

a) Cooper rating scale.

b) Cornell Aeronautical Lab. use of Cooper scale.

c) LSI rating scales used in UDOFTT simulator studies.

d) LSI rating scales used in attitude studies.

2) Comparative systems - pair people or units with respect to each other.

a) Comparison between pairs.

b) Card sorting technique.

3) Check lists - judgments by raters as to which of a series of descriptive terms either are or are not applicable to the units being evaluated.

4) Critical incidents - recording actual incidents as behaviors which are especially effective or ineffective in the accomplishment of the mission - Standing of unit is indicated by frequency of occurrence of reported incidents.

c) The sources of bias in ratings.

1) Halo effect.

2) Leniency error.

3) Error of central tendency.

4) Contrast error - tendency to rate in opposite direction on a dimension from how the raters see themselves.

5) Proximity error - tendency for ratings to be more related when made close to each other in time.

I. Individual vs. Crew Performance.

1) Question of what is "crew coordination" remains unanswered.

2) Groups have been studied from several conceptual viewpoints.

   a) Emphasis on group structure to understand implications of formal structure on group effectiveness.

   b) Group dynamics approach in context of a social systems with emphasis on the role, status and other factors which differentiate individuals within the group and influence group effectiveness.

   c) Group as a single man-machine system the effectiveness of its decisions and actions being shown by response adequacy, sequences of performance, and timeliness of behavior. These are determined by how crew members interact with each other and with equipment and the manner in which communications among crew members is achieved with regard to system output.

3) Group as a man-machine system is the approach to be taken in Human Factors field evaluation of Navy systems being considered here.

4) Crew performance must be regarded as more than the sum total of the individual performances.

5) Measures of crew performance.

   a) Synchronization of action.

   b) Response improvisation.

   c) Amount of time spent interacting - good crew should reduce individual interaction to a minimum so that more effort is devoted to the job and less to coordinating.

   d) Amount of communication - the less the communication the higher the degree of coordination.

   e) Freedom for interpersonal communications.

   f) Monitoring and/or making some responses for another crew member.

   g) Aiding in the detection of out-of-tolerance conditions.

   h) Sharing of risk activities among crew members.

J. Points at which to measure.

1) Points of interaction between man and machine at which inadequate performance significantly affects the accomplishment of the mission.

8. Procedural Steps in Assessment of Performance.

A. Conduct thorough analyses of the task.

B. Identify important and critical aspects of the task and the environment.

C. Define performance requirements of the task as appropriate.

D. Select test points and measures appropriate to the behavior to be evaluated.

E. Determine conditions under which measures will be taken.

F. Determine techniques for obtaining measurement data and for combining measures as appropriate.

G. Specify methods of data analysis.

9. Subjects or Operators upon which Data is Collected (Subject Sampling)

A. Sample size.

B. Sample composition with respect to experience and other factors and its relation to purposes of measurement and usefulness of the data.

10. The Subject Task and Environmental Conditions (Object Sampling)

A. Their relation to purposes of measurement.

B. Delineation of critical and relevant task and environmental parameters.

C. Specification and quantitification for manipulation as independent variables, for experimental control, or for assessment of the effect of their variation.

11. Data Collection and Treatment.

A. Experimental methods.

1) Single vs. multi-variate designs.

a) Single variable experimentation.

b) Multi-variate design.

(1) Each subject his own control.

(a) Appliciability for certain tasks but limitations of learning, practice and other experience in antecedent conditions of the experiment.

(b) Advantage is requirement for smaller total number of subjects.

(2) Independent groups.

(a) No interactive effect from cell to cell in the design.

(b) Subjects a problem particularly in field research.

B. Specific measures of performance.

1) Procedural tasks.

a) Time

b) Accuracy

2) Closed-loop tracking (Compensatory and pursuit)

a) Accuracy.

(1) Integrations of error ($\int e, \int |e|, \int e^2, \int (\bar{x}-x), \int (\bar{x}-x)^2$)

(Variability vs. average error)

(2) Number of crossings.

(3) Time on target.

(4) Frequency of catastrophic errors.

3) Crew coordination.

a) Overall gross measure of crew output in terms of:

(1) Time to accomplish task.

(2) Accuracy in accomplishing task.

b) Number of communications between members.

(1) Voice

(2) Other signals.

4) Decision Making Tasks.

a) Time

b) Accuracy

5) Perceptual and Motor Skills

a) Psychophysical measures.

C. Sample size and its relation to experimental method and data treatment.

1) Small number of subjects with many measurements per subjects for psychophysical data.

2) Small sample statistics - N = 30

3) Statistical significance vs practical significance.

4) Sample size and single variable experiments.

5) Sample size in multi-variate experiments.

a) Independent groups - N large enough to assume randomness or representativeness.

b) Subject his own control.

D. Parameteric - normal probability summary statistics.

1) Measures of central tendency - (Mean, Median, Mode)

2) Measures of deviation from standard (CE).

3) Measures of variability.

a) About a standard, (RMS, average deviation)

b) About mean (SD, Average deviation)

c) About median (Semi-interquartale and quartile range)

d) Range

e) Percentiles

f) Comparable scores (z scores, Stanines).

4) Measures of correlation.

a) Product-moment correlation

b) Percent argument (% = r^2)

5) Tests of reliability of differences.

a) t test (means, percentages)

b) chi square

c) Fiducial limits

6) Non-Parametric Statistics

a) Difference from parametric statistics.

b) Tests of reliability of differences.

(1) Wilcoxon signed ranks tests (related samples)

(2) Mann-Whitney U Test (independent samples)

c) Test of correlation.

(1) Spearman rank correlation (r_s)

(2) Kendall rank correlation coefficient (τ)

E. Presentation of Results.

1) Graphs

2) Bar Graphs (Histograms)

3) Significant differences in terms of probability.

DISTRIBUTION LIST FOR TECHNICAL REPORTS

Contract N00014-67-C-0315

No. of Copies	Addresses
1	Capt. E. C. Waller Project Manager, P-3, PMA 240 Naval Air Systems Command Department of the Navy Washington D. C. 20360
1	Mr. Meredith Munns Human Factors Engineering Office Naval Air Development Center Warminster, Pennsylvania 18974
1	Human Factors Section Electric Boat, General Dynamics Corp. Eastern Point Road Groton, Connecticut 06340
1	Dr. S. Roscoe Institute of Aviation University of Illinois Urbana, Illinois 61803
1	Dr. Carl Silver Drexel Institute of Technology College of Business Administration Philadelphia, Pennsylvania 19104
1	Whittenburg, Vaughan & Associates 4810 Beauregard Street Alexandria, Virginia 22302
1	Defense Contract Administration Services 500 South Ervay Street Dallas, Texas 75201
1	Scientific Advisory Team Commander ASW Forces, Atlantic Fleet Norfolk, Virginia 23511
1	Aerospace Medical Division Behavioral Science Laboratory Wright-Patterson AFB, Ohio 45433 Attn: Dr. J. M. Christensen (ASBSE)
1	Head, Human Factors Engineering Branch (5342) Naval Missile Center Point Mugu, California

No. of Copies	Addresses
1	Commanding Officer U. S. Naval Personnel Research Activity San Diego, California 92152
1	Head, Human Factors Division U. S. Navy Electronics Laboratory San Diego, California 92152
1	Commanding Officer Naval Air Technical Training Center Naval Air Station Memphis, Tennessee 38115
1	Director, Personnel Research Div. Chief of Naval Personnel (Pers A3) Department of the Navy Washington, D. C. 20370
1	Technical Director Personnel Research Laboratory Washington Navy Yard, Bldg 200 Washington, D. C. 20390
1	Dr. Robert R. Mackie Human Factors Research, Inc. 6780 Cortona Drive Santa Barbara Research Park Goleta, California 93107
1	Capt. J. E. Rasmussen, MSC, USN Chief of Naval Material (MAT 031M) Room 1017 Main Navy Building Washington, D. C. 20360
1	Dr. David Meister Human Factors Programs Bunker-Ramo Corp. 8433 Fallbrook Ave. Canoga Park, California 91304
1	Ronald A. Erickson Head, Human Factors Branch Code 3572 Naval Weapons Center China Lake, California 93555

No. of Copies	Addresses
20	Defense Documentation Center Cameron Station Alexandria, Virginia 22314
1	Commanding Officer Office of Naval Research Branch Office Box FPO New York 09510
6	Director U. S. Naval Research Laboratory Washington, D. C. 20390 Attn: Technical Information Div.
1	Head, Aviation Psychology Branch Bureau of Medicine and Surgery Department of the Navy Washiongton, D. C. 20390
1	Bureau of Medicine and Surgery Research Division Department of the Navy Washington, D. C. 20390
1	Aviation Psychology Division Naval Aerospace Medical Institute Naval Aerospace Medical Center Pensacola, Florida 32512
5	Chief of Naval Research (Code 455) Department of the Navy Washington, 20360 .
1	Director Office of Naval Research Branch Office 495 Summer Street Boston, Massachusetts 02210
1	Director Office of Naval Research Branch Office 1030 East Green Street Pasadena, California 91101
1	Director Office of Naval Research Branch Office 219 South Dearborn Street Chicago, Illinois 60604

No. of Copies	Addresses
2	Dr. K. Thomson Aero Medical Branch Service Test Division U. S. Naval Air Test Center Patuxent River, Maryland 20670
1	Dr. James J. Regan Head, Psychological Laboratory (Code 55) U. S. Naval Training Device Center Orlando, Florida 32813
1	Dr. J. A. Kraft Assistant Manager, Biotechnology Orgn. 55-60, Bldg. 151 Lockheed Missile and Space Co. Sunnyvale, California 94088
1	Dr. John J. Collins Office of the Chief of Naval Operations Pentagon Washington, D. C. 20350
1	CDR John P. Charles Office of the Chief of Naval Op-701H Department of the Navy Washington, D. C. 20350
1	Office of Naval Research Resident Representative Rooms 2507-08, Main Building P. O. Box 7938 Austin, Texas 78712
1	Commandant of the Marine Corps G-1 Division (Code A01B) Navy Department Washington, D. C. 20380
1	U. S. Naval Medical Research Lab. U. S. Naval Submarine Medical Center New London, Connecticut 06320 Attn: Dr. G. O. Moeller
1	Technical Director U. S. Army Human Engineering Lab. Aberdeen Proving Ground Aberdeen, Maryland 21005

No. of Copies	Addresses
1	Dr. Robert R. Mackie Human Factors Research, Inc. 6780 Cortona Drive Santa Barbara Research Park Goleta, California 93107
1	Capt. J. E. Rasmussen, MSC, USN Chief of Naval Material (Mat 031) Room 1017 Main Navy Building Washington, D. C. 20360
1	Dr. David Meister Human Factors Programs Bunker-Ramo Corp. 8433 Fallbrook Ave. Canoga Park, California 91304
1	Ronald A. Erickson Head, Human Factors Branch Code 3572 Naval Weapons Center China Lake, California 93555
1	Dunlap and Associates 1454 Cloverfield Blvd. Santa Monica, California 90404
2	CDR. A. E. McMichael Human Factors Engineering Office Naval Air Development Center, Johnsville Warminster, Pennsylvania 18974
1	Dr. Arthur I. Siegel Applied Psychological Services Science Center 404 East Lancaster Avenue Wayne, Pennsylvania 19087
2	Commander Operational Test and Evaluation Force U. S. Naval Base Norfolk, Virginia 23511
1	Lt. J. Funaro, MSC, USN Air 5313 Naval Air Systems Command Washington, D. C. 20360

No. of Copies	Addresses
1	Dr. Heber Moore Office of Naval Material (Code 0325) Department of the Navy Washington, D. C. 20360
1	Director, Aeronautics Programs Department of the Navy Office of Naval Research (Code 461) Washington, D. C. 20360
1	Scientific Advisory Team Commander, ASW Forces, U.S. Atlantic Fleet, Code 71 Norfolk, Virginia 23511
1	Aeronautics Programs, Code 461 Aircraft Instrumentation Project Officer Office of Naval Research Washington, D. C. 20360
	Dr. A. L. Slafkosky Scientific Advisor Commandant of the Marine Corps (Code AX) Washington, D. C. 20380

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Life Sciences, Inc. 7305 Grapevine Highway Fort Worth, Texas 76118	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

HUMAN FACTORS IN FIELD TESTING

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report March 1, 1967 - February 28, 1969

5. AUTHOR(S) (First name, middle initial, last name)

W. G. Matheny

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
N00014-67-C-0315-P001	76	92

8a. CONTRACT OR GRANT NO. b. PROJECT NO. 006-09-01 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) LS-TR-69-1 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; Its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
Technical Report as a part of an ongoing project	Engineering Psychology Programs Psychological Sciences Division Office of Naval Research, Washington, D.C.

13. ABSTRACT

This study examined the problem of human factors field evaluation within the Navy for the purpose of formulating recommendations for improved evaluative methods and techniques. Conclusions and recommendations were drawn from reviews of relevant general literature and Navy documents, interviews and discussions with individuals experienced in human factors evaluation, and an examination of the design and evaluation procedures carried out during the development and test of a specific Navy aircraft system.

Conclusions and recommendations are enumerated in Section 6.0 of the report along with references to the relevant supporting Section (s) within the body of the report.

Briefly, it was concluded that human factors evaluation does not receive emphasis or support comparable to that given equipment evaluation or commensurate with the importance of the human operator to the successful functioning of the system. Much more definitive and timely information must be provided the human factors field evaluator, evaluations must be more mission oriented rather than cockpit centered, the role of the mock-up inspection needs redefinition, assignment of trained Navy human factors personnel to advise and assist the contractor during development is recommended as is assignment of contractor human factors personnel during field evaluations, close cooperation between human factors and equipment design evaluation personnel during field evaluation will greatly increase the effectivenes[illegible] [illegible]e evaluation, and a short intensive training course in human factors eval[illegible] problems and methods is recommended for Navy personnel assigned to sys[illegible] [illegible]luation.

DD FORM 1 NOV 65 1473

Unclassified

Security Classification

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Design Mission						
Evaluation						
Field Test						
Human Factors Training						
Mathematical Models						
Measurement						
Mock-Up Inspection						
Performance						
Simulation						
System Analysis						